

**Università degli Studi di Napoli
“FEDERICO II”**

DIPARTIMENTO DI FARMACIA



**Dottorato di ricerca in “Scienze del farmaco”
XXVI CICLO**

**Metabolomic NMR fingerprinting:
an exploratory and predictive tool**

Tutor:

Chiar.mo Prof. Ettore Novellino

Candidate:

Ilaria Lauri

Co-tutor:

Assoc. Prof. Antonio Randazzo

PhD Coordinator: Prof. Maria Valeria D'Auria

ABSTRACT

Metabolomics is the comprehensive assessment of low molecular weight organic metabolites within biological system. The identification and characterization of several chemical species, or metabolic *fingerprinting*, is an emergent approach in metabolomics field that provides a valuable “snapshot” of metabolic profiles. This approach is finding an increasing number of applications in many areas including cancer research, drug discovery and food science. The combined use of NMR spectroscopy, data pre-processing tools, and multivariate statistical data analysis allows to go through into the metabolite “signature” of various samples. The PhD project focused on the use of metabolic NMR fingerprinting as an exploratory and predictive tool. The first study tested the potentiality of Nuclear Magnetic Resonance spectroscopy as “magnetic tongue”: the NMR metabolomic signature can differentiate canned tomato samples, on the basis of their chemical composition and can be correlate to the sensory descriptors. *Orthogonal projection to latent structures* (OPLS) models were performed to demonstrate the NMR potentiality to predict the sensory descriptors. The second study showed the applicability of this methodology to measure and to predict sensory descriptors in extra-virgin olive oil. The third study demonstrated that the combined use of NMR spectroscopy and chemometrics can provide a cocaine seizures profiling, improving police investigation strategies about the cocaine trafficking routes and distribution network.

The last study of the PhD project, performed at the University of Copenhagen (Department of Food Science, Faculty of Life Science), concerns the analysis of metabolomic profiles of human colon cancer cell lines. The study includes a development of the experimental protocol for an efficient harvesting, quenching and extraction of cellular metabolites of HTC-116 human adherent cancer cell lines in order to analyze colon cancer cells metabolome and to understand *in vitro* actions of novel anticancer drugs.

LIST OF PUBLICATIONS

Paper 1

Malmendal A., Amoresano C., Trotta R., **Lauri I.**, De Tito S., Novellino E., Randazzo A. NMR Spectrometers as “Magnetic Tongues”: Prediction of Sensory Descriptors in Canned Tomatoes. *Journal of Agricultural and Food Chemistry*, 2011, 59 (20), 10831-1835.

Paper 2

Lauri I., Pagano B., Malmendal A., Sacchi R., Novellino E., Randazzo A. Application of “magnetic tongue” to the sensory evaluation of extra virgin olive oil. *Food Chemistry*, 2013, 140 (4), 692-699.

Paper 3

Pagano B., **Lauri I.**, De Tito S., Persico G., Chini M.G., Malmendal A., Novellino E., Randazzo A. Use of NMR in profiling of cocaine seizures. *Forensic Science International*, 2013, 231 (1-3), 120-124.

Other publications

Trotta R., De Tito S., **Lauri I.**, La Pietra V., Marinelli L., Cosconati S., Martino L., Conte M.R., Mayol L., Novellino E., Randazzo A. A more detailed picture of the interactions between virtual screening-derived hits and the DNA G-quadruplex: NMR, molecular modelling and ITC studies. *Biochimie*, 2011, 93 (8), 1280-1287.

Cosconati S., Rizzo A., Trotta R., Pagano B., Iachettini S., De Tito S., **Lauri I.**, Fotticchia I., Giustiniano M., Marinelli L., Giancola C., Novellino E., Biroccio A., Randazzo A. Shooting for selective druglike G-quadruplex binders: Evidence for telomeric DNA damage and tumor cell death. *Journal of Medicinal Chemistry*, 2012, 55 (22), 9785-9792.

Pagano B., Cosconati S., Gabelica V., Petraccone L., De Tito S., Marinelli L., La Pietra V., di Leva F.S., **Lauri I.**, Trotta R., Novellino E., Giancola C., Randazzo A. State-of-the-Art Methodologies for the Discovery and Characterization of DNA G-Quadruplex Binders. *Current Pharmaceutical Design*, 2012, 18, 1880-1899.

LIST OF ABBREVIATIONS

ADP	Adenosine diphosphate
AMBIC	Ammonium bicarbonate
AMP	Adenosine monophosphate
ATP	Adenosine Triphosphate
COW	Correlation optimized warping
CPMG	Carr-Purcell-Meiboom-Gill (NMR pulse sequence)
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
DSS	Sodium 2,2-dimethyl-2-silapentane-5-sulfonate
EVOO	Extra-virgin olive oil
FID	Free induction decay
GC-MS	Gas chromatography-mass spectrometry
G4-DNA	G-tetrads or G- <i>quadruplex</i> DNA
HCA	Hierarchical Cluster Analysis
HMDB	Human Metabolome Database
KOH	Potassium hydroxide
LC-MS	Liquid chromatography-mass spectrometry
NMR	Nuclear Magnetic Resonance
MDMA	Methylenedioxy- N-methylamphetamine
NOESY	Nuclear Overhauser effect spectroscopy
1D-NOESY	One-dimensional-Nuclear Overhauser effect spectroscopy
OPLS	Orthogonal projection to latent structures
PBS	Phosphate buffered saline
PC	Principal component
PCA	Principal component analysis
PLS	Partial Least Squares regression
PLS-DA	Partial Least Squares-Discriminant Analysis
QDA	Quantitative descriptive analysis
RHPS4	3,11-Difluoro-6,8,13-trimethyl-8H-quino[4,3,2-kl]acridinium methosulfate
RNA	Ribonucleic acid
STOCSY	Statistical Total Correlation Spectroscopy
Tel24	Human telomeric G-quadruplex structure
TMS	Tetramethylsilane
TSP	(trimethylsilyl) propanoic-2,2,3,3-d ₄ acid

Contents

List of publication	I
List of abbreviations	II
1. Introduction	1
1.1 Aim of the project	2
1.2 Thesis outline	3
2. What is metabolomics?	4
2.1 System Biology: Metabolomics and Metabonomics	4
2.2 Designing a metabolomic experiment: Targeted and Untargeted metabolomics	7
2.3 Analytical platform: ¹ H NMR	9
2.4 Data pre-processing	11
2.4.1 <i>Alignment</i>	11
2.4.2 <i>Binning</i>	12
2.4.3 <i>Normalization</i>	12
2.5 Multivariate Data Analysis	13
2.5.1 <i>Centering and scaling</i>	14
2.5.2 <i>Principal Component Analysis (PCA)</i>	15
2.5.3 <i>Hierarchical Cluster Analysis (HCA)</i>	17
2.5.4 <i>Partial Least Squares-Discriminant Analysis (PLS-DA)</i>	18
2.6 References	20
3. Case Studies	24
3.1 NMR Metabolomic <i>fingerprinting</i> : a tool for prediction of sensory descriptors in canned tomatoes	24
3.1.1 <i>Project workflow</i>	26
3.1.2 <i>Methods and results discussion</i>	27
3.2 Application of “magnetic tongue” to the sensory evaluation of extra virgin olive oil	33
3.2.1 <i>Methods and results discussion</i>	34
3.3 NMR profiling of cocaine seizures	40
3.3.1 <i>Methods and results discussion</i>	41
3.4 Conclusions	47
3.5 References	48

4. NMR Metabolomics Of Human Cancer Cell Lines	51
4.1 Introduction	51
4.2 Methodology of sample preparation and extraction	53
4.3 Anticancer drugs and DNA <i>G-quadruplex</i> binders.....	55
4.4 Experimental design.....	62
4.5 Materials and methods	64
4.6 Experimental protocol optimization for NMR metabolomic analysis of HTC-116 cell lines	67
4.7 NMR sample preparation	70
4.8 ¹ H NMR Data acquisition.....	70
4.9 NMR Data Analysis and selection of experimental protocol.....	72
4.10 Pre-Processing of the NMR spectra and Multivariate Data Analysis.....	74
4.11 Metabolites identification.....	84
4.12 Conclusions and future perspectives	87
4.13 References	89

1 INTRODUCTION

*Data does not equal information; information does not equal knowledge; and,
most importantly of all, knowledge does not equal wisdom.*

*We have oceans of data, rivers of information, small puddles of knowledge,
and the odd drop of wisdom.*

Henry Nix, 1990

Data, information and knowledge represent the keywords in metabolomics field. Metabolomics – the dynamic portrait of the metabolic status of living system – is a strongly developing field as evident from the exponentially growing number of papers. Over the past decades, metabolomics approaches have been widely adopted in many areas, such as in cancer research, in drug discovery and in food science. Studies of the metabolome include the analysis of a large range of several chemical species, providing challenges to all analytical technologies employed in metabolomics strategies. Nowadays, proton Nuclear Magnetic Resonance (^1H -NMR), Gas Chromatography-Mass Spectrometry (GC-MS) and Liquid Chromatography-Mass Spectrometry (LC-MS) are well-established powerful analytical methods for generating metabolomics profiles. In particular, developments in NMR offer distinct advantages for performing untargeted metabolomic studies because of its specificity and quantitative reproducibility.

1.1 Focus and aim of the project

In this thesis, a *metabolomic fingerprinting approach* was used in order to differentiate canned tomato samples, on the basis of their chemical composition and then to correlate the *fingerprints* to the sensory descriptors. To evaluate the applicability of this methodology, the same strategies and study design were applied to measure and predict sensory descriptors in extra-virgin olive oil.

The primary aim of this PhD projects was to test the potentiality of Nuclear Magnetic Resonance spectroscopy as “magnetic tongue” to measure sensory descriptors of food samples, applying the metabolomic fingerprinting approach. We will try to demonstrate that NMR could be a powerful predictive tool for the characterization of sensory features.

The same approach was applied in forensic science field in order to identify the chemical “fingerprint” of cocaine samples and to test the NMR potentialities, in combination with multivariate data analysis, to improve and support police investigations.

During last PhD year at the University of Copenhagen, I focused my attention on the analysis of metabolic profiles of human colon cancer cell lines – using NMR spectroscopy-based metabolomics – in order to provide important information on *in vitro* actions of drugs, pointing in their rapid incorporation into novel therapeutic settings. This study aimed to develop an optimized experimental protocol for NMR metabolomics study of HTC 116 human colon cancer cell lines.

1.2 Thesis outline

The thesis is subdivided as follow:

Chapter 2: presents a general description of metabolomics science focusing the attention on the untargeted approach; the analytical platform used and the multivariate data models performed are briefly described.

Chapter 3: presents an overview of the results and discussion from paper I, II and II.

Chapter 4: presents unpublished data about the project that I carried out at the University of Copenhagen, Food Science department, as PhD visiting student, under the supervision of Assoc. Prof. Francesco Savorani and Prof. Søren Balling Engelsen.

2 WHAT IS METABOLOMICS?

2.1 System Biology: Metabolomics and Metabonomics

The term *metabolomics* was introduced for the first time in the early 2000s by Oliver Fiehn as “the identification and quantification of all metabolites” in a biological system [1].

Few years before, in the 1999, Jeremy Nicholson and colleagues formally defined a similar term *metabonomics* as “the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification”[2], but this concept was born with the first simultaneous analysis of metabolites present in biological fluids through ^1H NMR spectroscopy in the 1980s [3].

Therefore historically, the metabonomics approach was one of the first methods to apply the scope of systems biology to studies of metabolism.

There has been some disagreement over the exact differences between the two terms *metabolomics* and *metabonomics*. The distinction is mainly philosophical, rather than technical; although there are some differences in concept, in practice, the analytical and modeling procedures are the same and the two terms are often used interchangeably by scientists and organizations [4].

Metabolomics represents the dynamic portrait of the metabolome, that is the collection of metabolites (low-molecular-weight molecules-intermediates and the end products of metabolic reactions) belonging to many classes of

compounds, such as amino acids, organic acids, lipids, nucleotides, etc, within a biological system.

The estimated size of the metabolome is large: *S. cerevisiae* approximately 600 metabolites [5], plant kingdom up to 200 000 metabolites [6], and analysis of the human metabolome reveals greater complexity. In particular, as released by The Human Metabolome Database (HMDB) (www.hmdb.ca), the number of annotated metabolite entries has grown from 6500 to more than 40000, including “detected” metabolites (with experimental confirmation of their existence) and “expected” metabolites (for which biochemical pathways are known but the compound has yet to be detected in the body) [7].

Metabolomics is the newest “omics” science. The suffix “-omics” has been added to the names of many fields to denote studies undertaken on a very large-scale data collection and analysis, i.e. measuring/profiling a large number of variables simultaneously.

Metabolomics can represent the final product of the “omics” cascade of the interactions between genes, proteins and metabolites with cellular environment, adding the end point of the building blocks of System biology (Genomics, Transcriptomics, Proteomics) (Fig.1). Therefore changes in the metabolome are the ultimate answer of an organism to genetic alterations, disease, or environmental influences [8].

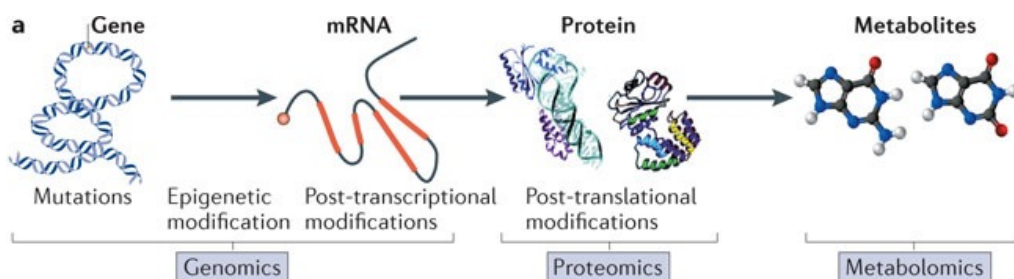


Fig.1 The “omics” cascade (modified from Gary J. Patti et al., 2012).

One problem with the metabolomics approach is given by different timescales of “omics” events (gene expression, protein expression and metabolites generation). The transcript machinery responds fast to an external stimulus (seconds to minutes), the proteins may be expressed within minutes to hours and metabolites vary significantly during the day [9]. Therefore, metabolomics is more time sensitive than the other “omics”. Since environmental and lifestyle factors influence metabolism, it’s difficult to separate these effects from gene-related effects [4]. Metabolomics overcomes these problems by monitoring the global outcome of all the influencing factors in a holistic approach, without making assumptions about the effect of any single contribution to that outcome [4].

2.2 Designing a metabolomic experiment: Targeted and Untargeted metabolomics

The first step in performing metabolomics is to decide which type of approach needs to use on the strength of the aim of study and the kind of information to obtain from the metabolomic analysis (Fig.2). In some cases, it may be of interest to examine a defined set of metabolites by using a *targeted approach*. On the other hand, an *untargeted or global approach* may be taken in which as many metabolites as possible are measured and compared between samples [10].

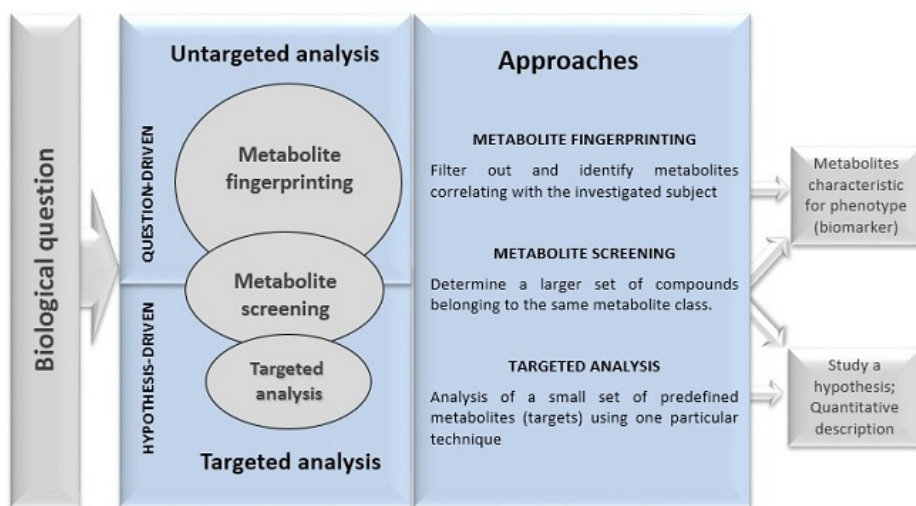


Fig.2 Strategies for metabolomic investigations.

Targeted analysis focuses on the absolute quantification of a small set of metabolites, identified in advance, that are highly related to a specific pathway or intersecting pathways, after an appropriate sample preparation to

separate metabolites from sample matrix [11].

This kind of analysis – in literature also known as *targeted profiling* or *quantitative metabolomics* – is characterized as an hypothesis-driven approach rather than an hypothesis-generating. However, targeted analysis is not used in this PhD study and will not be discussed further.

In contrast, *Metabolite profiling*, probably the mostly applied untargeted approach, involves rapid analysis, often not quantitative, of a large number of different metabolites with the objective of identifying a specific metabolite profile that characterizes a given sample. Therefore, untargeted approach applies different analytical platforms such as NMR, GC-MS, LC-MS. This approach can be subdivided into *metabolomic fingerprinting* and *metabolomic footprinting* [11] (see Chapter 4).

Metabolomic fingerprinting is a global analysis of crude samples or sample extracts with minimal preparation, for sample classification or screening of samples. Usually in this approach not all metabolites must be identified and quantified, since its attention is focused on giving a rapid snapshot, or fingerprint, of phenotypes [12].

This PhD thesis describes metabolomic fingerprinting as an exploratory and predictive tool for investigation and discrimination of samples from different origins by using NMR as analytical platform.

2.3 Analytical platform: ^1H NMR

Nuclear magnetic resonance (NMR) spectroscopy is a valuable analytical technique in modern chemical research. It is a non-destructive and non-invasive analytical method that requires minimal or no sample preparation and doesn't alter the sample composition. NMR is frequently used in the qualitative and quantitative analysis of small organic compounds as well as structure analysis and interpretation. Thus, ^1H NMR represents a powerful tool in numerous applications, especially in metabolomics research and also in food science and technology field. The main disadvantage of NMR spectroscopy, if compared to MS, is the low sensitivity, but with current ^1H NMR spectroscopy instruments with higher magnetic field strength and cool cryogenic probes, this method is widely used in non-targeted metabolomics.

In this PhD study ^1H NMR spectra were acquired at 25°C with a 700 MHz Varian Unity Inova spectrometer using a 5mm $^1\text{H}\{^{13}\text{C}/^{15}\text{N}\}$ triple resonance probe. The ^1H -NMR measurements were carried out with 1000 transients and 32 K complex data point. In order to retrieve quantitative information, the recycle time was set to 5 s, and a 45° pulse angle was used.

The spectra were processed using iNMR software (www.inmr.net). An exponential line broadening of 0.5 Hz was applied to the free-induction decay prior to Fourier transformation. All spectra were referenced relative to external sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS), manually phased and automatically baseline corrected (Fig.3).

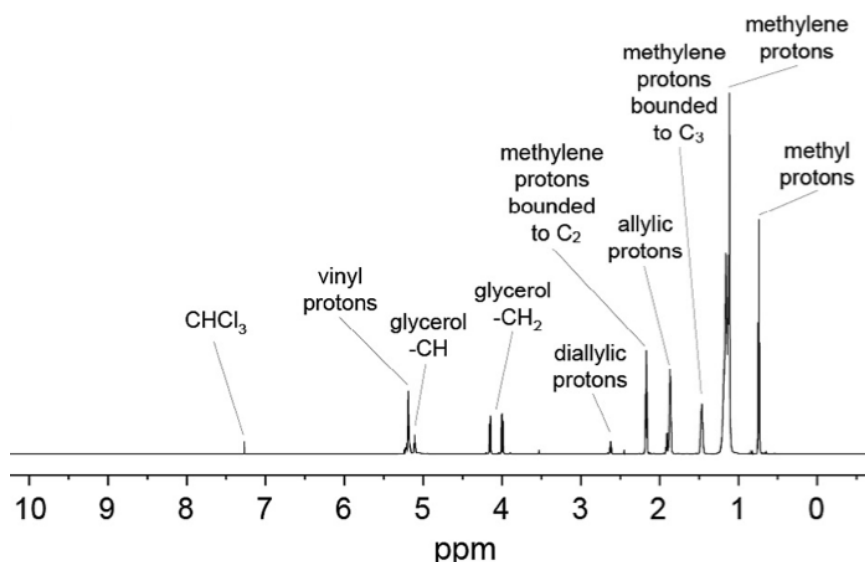


Fig.3 NMR spectrum of a representative sample of extra-virgin olive oil. (Paper II)

In the last project carried out at University of Copenhagen (Department of Food Science), Carr-Purcell-Meiboom-Gill (CPMG) and 1D Nuclear Overhauser Effect Spectroscopy (NOESY) experiments were performed for acquiring NMR spectra of hydrophilic/lipophilic cancer cell extracts and the growth culture media.

In this section the differences between two experiments will be described. A detailed description of the data acquisition is reported in the *Chapter 4*.

The ^1H NMR spectra of aqueous extracts are measured using a specified water suppression pulse sequence such as 1D NOESY-presat, which employs the first increment of a NOESY pulse sequence with water irradiation during the relaxation delay and also during the mixing time. This experiment reduces contributions from regions of the active volume that experience an incomplete 90°C pulse, thus reducing the residual water resonance [13].

The CPMG-presat pulse sequence is used in order to discriminate between metabolites of low molecular weight (typically <1000 Da) and macromolecular species in the sample, such as proteins or lipoproteins. Macromolecules produce broad resonances, due to reduced rotational diffusion and short T_2 relaxation times, which confound spectral interpretation. Commonly and in this study (see *chapter 4*) T_2 editing via the CPMG experiment is used to reduce the contribution of high molecular weight species in the resulting spectra [14].

2.4 Data pre-processing

After acquiring the NMR data, it is important to use an appropriate data pretreatment methods prior starting data analysis. Different data preprocessing steps [15] are applied in order to generate “clean” data, that will represent the input for data analysis.

In *all investigations*, the NMR spectral regions including the residual solvent signal, or signals strongly affected by it, were removed in order not to compromise the analysis. The high and low-field ends of the spectrum were also removed because they include only noise, which does not provide relevant information.

In my PhD study pre-processing methods were applied: spectra alignment, binning, normalization.

2.4.1. Alignment

In metabolomic NMR data analysis solving signal alignment problems could be a powerful tool for multivariate exploratory investigations aimed at biomarker profiling or pattern recognition studies.

In the *paper II* and in the project “*NMR metabolomic of cancer cell lines*” (*chapter 4*), the spectra were corrected for misalignments in chemical shift due to pH sensitive signals using the *interval-based icoshift algorithm* (the Matlab code including documentation can be downloaded from www.models.life.ku.dk) [16].

In the *paper I* the spectra were aligned by the segmented warping method *correlation optimized warping* (COW) using $m_p = 50$ and $n_p = 2$ [17].

2.4.2. Binning

The standard approach to solve problems of unmanageable dimensionality of the data and the inter-individual differences in peak locations is the division of each spectrum in equally sized bins, integration of the intensity values in each bin, and annotation of this value to the bin [15].

In the *paper I, II and III* data reduction was accomplished by dividing the spectrum into bins (0.01 ppm-*paper I*, 0.005 ppm-*paper II and III*) over which the signal was integrated to obtain the signal intensity.

In the project “*NMR metabolomic of cancer cell lines*” the spectra alignment, using *icoshift* algorithm, solved peak misalignment problems. For a more detailed description see *Chapter 4*.

2.4.3. Normalization

In metabolomics, data are usually presented as a table where each row relates to a given sample and each column corresponds to individual spectral peak intensities. NMR spectroscopy is a technique that allows quantitative analysis. Thus the signal intensities should be correlated to the metabolites concentration. Unfortunately, technical variations originated from sampling,

general instrumental sensitivity effects can affect the signal intensities.

Sample normalization is a row operation that is applied to the data from each sample and comprises methods to make the data from all samples directly comparable with each other [18].

In *paper I* the integrals were normalized to a total intensity to suppress trivial separation based on variations in the amount of sample.

In *paper II*, the integrals were normalized to the integral of the triplet at δ_H 0.86 ppm (CH₃ of triacylglycerols) and in *paper III* using the cocaine triplet at δ_H 7.65 ppm (H₄₀).

2.5 Multivariate Data Analysis

First of all, why *multivariate* data analysis?

To understand the world around us we need to measure many variables, many properties of the systems we investigate. Data collected in science and technology fields are multivariate, with multiple variables on multiple samples.

Multivariate data, accurately measured on selected observations and variables, contain much more information than univariate data. In order to obtain insight into the system studied, the first step in analyzing set of data is the raw data exploration. Plotting the data could be a good approach in order to visualize different features of the data: phase- baseline correction problems (in the case of NMR data sets), sample with peculiar deviations of the data, outlying samples and expected biomarkers [19].

However, in order to extract hidden information and obtain relations between variables is not enough to just look at the raw data.

Explorative unsupervised and classification multivariate methods

(chemometrics) are useful to investigate the data and identify meaningful patterns in the data.

Prior to multivariate data analysis, data are often pre-treated, in order to transform the data into a form suitable for analysis. In this section centering and scaling of data are described.

2.5.1. Centering and scaling

In NMR metabolomics data, the feature intensities vary for metabolite signals; not always the abundant compounds are more important than the lower ones. Moreover, using Principal Component Analysis (PCA) which is a maximum variance projection method, a variable with a large variance is more likely to be expressed in the modeling than a low-variance variable. , For instance, NMR spectral regions including intense signals will have more importance in the model than the regions with lower ones. Thus using the combination of centering and scaling is recommended.

Centering removes the offset from the data between high and low abundant metabolites. In my studies mean-centering was used in order to improve the interpretability of the models.

$$\tilde{x}_{ij} = x_{ij} - \bar{x}_i$$

With mean-centering the average value of each variable is calculated and then subtracted from the data.

Among scaling types, Pareto scaling was applied in all papers prior chemometric analysis.

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{\sqrt{s_i}}$$

It gives each variable a variance numerically equal to its initial standard deviation, reducing the relative importance of large values, but keeping data structure partially intact [20].

In this PhD study some chemometric tools were used to extract information from NMR data and they are briefly described below.

2.5.2. Principal Component Analysis (PCA)

PCA was first formulated in statistics by Pearson, who described the analysis as finding lines and planes of closest fit to systems of points in space [21]. PCA is a multivariate projection method designed for an exploratory (unsupervised) data analysis in order to display the systematic variation in a data matrix, extracting hidden information.

Statistically, PCA finds lines, planes, hyperplanes in the K-dimensional space that approximate the data as well as possible in the least squares sense. It finds the directions in multivariate space that represent the largest sources of variations, the so called *principal components* (PC's).

Graphically, PCA models build a score and a loading plot. The score plot shows how the observations are projected onto this planes, while the loading plot displays the relationships among the variables (Fig.4).

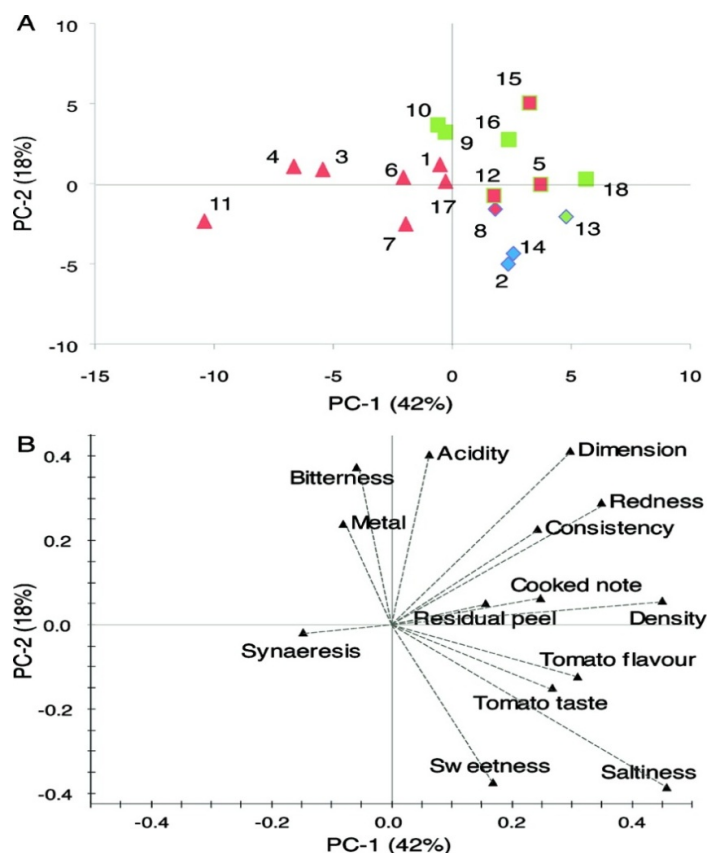


Fig.4 Score (A) and loading (B) plots of the PCA performed on sensory data.

In metabolomic studies, PCA is used as a tool for an exploratory approach to unknown data. It allows to find peculiar data trends, outliers and similarities or dissimilarities among samples.

This unsupervised data analysis was applied to NMR data from all papers, including the unpublished data in *Chapter 4*.

2.5.3. Hierarchical Cluster Analysis (HCA)

Hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters [22]. The HCA algorithm connects objects to form clusters based on their distance. The inputs required are similarity measures or data from which similarities can be computed using different distance functions. The main property of HCA is to highlight grouping of samples on the basis of similarities or distances (dissimilarities) with the general idea that objects are more related to nearby objects than to objects farther away.

The results of hierarchical clustering are presented in a dendrogram, in which the y-axis marks the distance among clusters, while the objects are placed along the x-axis (Fig.5).

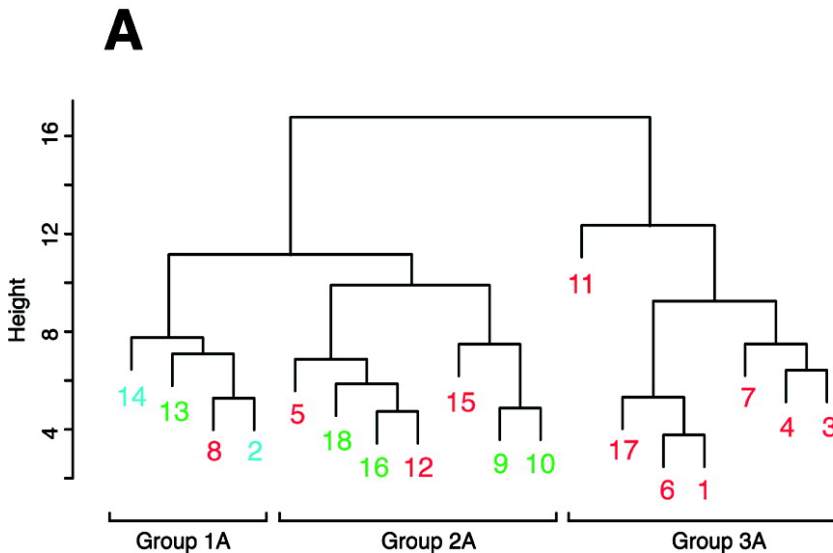


Fig.5 HCA dendrogram showing the similarities between products based on QDA (Paper I)

In the *paper I*, HCA was carried out using the Euclidean distance between the PLS scores for each canned tomato sample, in order to group samples with similar sensory features.

In *paper II*, HCA was performed on quantitative descriptive analysis (QDA) means, in order to group extra virgin olive oil samples sharing similar sensory features.

In *paper III*, it was carried out by using the Ward clustering method [22], in order to group cocaine samples with different periods and areas of the cocaine consignments.

2.5.4. Partial Least Squares-Discriminant Analysis (PLS-DA)

In metabolomics, Partial Least Squares-Discriminant Analysis (PLS-DA) is commonly used for classification in multivariate data analysis.

PLS-DA is a regression model that describes the maximum separation between pre-defined classes [23]. The objective of PLS-DA is to find a model that separates classes of samples on the basis of their variables.

In PLS-DA, the X-matrix consists of sample variables (features) and, in order to encode a class identity, the response Y is a matrix of dummy variables. This artificial matrix is defined by creating a vector for each class, where each Y-variable has value of 0 for all samples in the first class, of 1 for all samples in the second class and so on. Then, the PLS-DA algorithm can find a discriminant plane in X-space in which the projected samples are well separated according to class.

In *paper III*, PLS-DA was performed using Simca-P 12.0 (Umetrics) to explore the ability of the NMR data to discriminate between classes of samples. The quality of the models was described by R^2 and Q^2 values. R^2 is defined as the proportion of variance in the data explained by the models and indicates the goodness of fit. Q^2 is defined as the proportion of variance in the data predictable by the model and indicates predictability [24]. The model validation was performed using the permutation test [25], in which a total of 400 models were calculated by randomizing the order of Y variables in the corresponding PLS-DA models.

After the interpretation of multivariate data analysis models, the last and more challenging step in metabolomic studies is the NMR assignment of the selected biomarkers. In the next chapter more detailed results are reported for each case study.

2.6 REFERENCES:

- [1] Fiehn O., Metabolomics-the link between genotypes and phenotypes. *Plant Molecular Biology*, **2002**, 48, (1-2), 155-171.
- [2] Nicholson J. K., Lindon J. C., Holmes E., 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*, **1999**, 29, (11), 1181-1189.
- [3] Wilson I.D., Wade K.E., Nicholson J.K., Analysis of biological fluids by high-field nuclear magnetic resonance spectroscopy. *Trends in Analytical Chemistry*, **1989**, 8, (10), 368-374.
- [4] Nicholson J.K., Lindon J.C., Sistem biology: Metabonomics. *Nature*, **2008**, 455, (7216), 1054-1056.
- [5] J. Forster I. Famili P. Fu, B. O. Palsson and J. Nielsen, Genome-Scale Reconstruction of the *Saccharomyces cerevisiae* Metabolic Network. *Genome Res.*, **2003**, 13, 244–253.
- [6] O. Fiehn, Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp. Funct. Genomics*, **2001**, 2, 155–168.
- [7] David S. Wishart *et al.* HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic Acids Research*, **2012**, 1-7.
- [8] Rochfort S. Metabolomics reviewed: a new “omics” platform technology for systems biology and implications for natural products research. *J.Nat Prod*, **2005**, 68, 1813-1820.

- [9] Kussmann M., Raymond F. Affolter M. OMICS-driven biomarker discovery in nutrition and health. *Journal of Biotechnology*, **2006**, 124, 758–787.
- [10] Patti G.J., Yanes O., Siuzdak G. Innovation: Metabolomics: the apogee of the trilogy. *Nature Reviews Molecular Cell Biology*, **2012**, 13, 263-269.
- [11] Claudino W.M., Quattrone A., Biganzoli L., Pestrin M., Bertini I., Di Leo A. Metabolomics: Available Results, Current Research Projects in Breast Cancer, and Future Applications. *Journal of Clinical Oncology*, **2007**, 25, 2840-2846.
- [12] Dunn W.B. Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Phys. Biol.*, **2008**, 5.
- [13] Neuhaus D., Ismail I.M., Chung C.W. “FLIPSY” – A New Solvent-Suppression Sequence for Nonexchanging Solutes Offering Improved Integral Accuracy Relative to 1D NOESY. *J Mag Res*, **1996**, A, 118, 2, 256–263.
- [14] Meiboom S., Gill D. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Rev. Sci. Instrum.*, **1958**, 29, 688.
- [15] De Meyer T., Sinnaeve D., Van Gasse B., Tsiorkova E., Rietzschel E.R., De Buyzere M.L., Gillebert T.C., Bekaert S., Martins J.C., Crieckinge W.V. NMR-Based Characterization of Metabolic Alterations in Hypertension Using an Adaptive, Intelligent Binning Algorithm. *Anal. Chem.*, **2008**, 80, 3783–3790.

- [16] Savorani F., Tomasi G., Engelsen S.B. *icoshift*: A versatile tool for the rapid alignment of 1D NMR spectra. *J Mag Res*, **2010**, 202, 190–202.
- [17] Tomasi, G., van den Berg, F., Andersson, C. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *J. Chemom.*, **2004**, 18, 231–241.
- [18] Craig A., Cloarec O., Holmes E., Nicholson J.K., Lindon J.C. Scaling and Normalization Effects in NMR Spectroscopic Metabonomic Data Sets *Anal. Chem.* **2006**, 78, 2262-2267.
- [19] Eriksson, L., Johansson, E., Kettaneh-Wold, N., & Wold, S. (2006). Multi- and megavariate data analysis – Part I: Basic Principles and Applications. Umetrics.
- [20] van den Berg R.A, Hoefsloot Huub CJ, Westerhuis J.A, Smilde A.K, van der Werf M. J. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, **2006**, 7:142.
- [21] Pearson, K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, **1901**, 2,559-572.
- [22] Ward J.H., Hierarchical grouping to optimize an objective function, *J. Am. Stat. Assoc.*, **1963**, 48, 236–244.
- [23] Pérez-Enciso M., Tenenhaus M. Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach. *Human Genetics*, **2003**, 112, 5-6, 581-592.
- [24] Szymanska E., Saccenti E., Smilde A.K., Westerhuis J.A. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics*, **2012**, 8, 3–16.

- [25] Westerhuis J.A., Hoefsloot H.C.J., Smit S., Vis D.J., Smilde A.K., van Velzen E.J.J., van Duijnhoven J.P.M., van Dorsten F.A., Assessment of PLSDA cross validation. *Metabolomics*, **2008**, 4, 81–89.

3 CASE STUDIES

3.1 NMR Metabolomic *fingerprinting*: a tool for prediction of sensory descriptors in canned tomatoes

Sensory perception appears so deceptively easy and straightforward that it may seem to be a simple detection process merely “capturing” the environmental signals and feeding them into the brain. But the perception of odor and flavor of food is a complex physiological and also psychological process that cannot be explained by simple models.

The sensory food impression is mainly determined by the chemical senses of taste and smell. Both are detected through sensory cells of the tongue (taste) and of the nasal cavity (smell) [1], on the base of different molecules or ions.

Nowadays understanding consumer’s expectations, habits, and preferences is really important especially for a food company in order to ensure the product success on the market. Of course brand, label, packaging, advertising are critical factors in product choice matter, but the success is tightly connected to the products’ features. Unfortunately consumers are not able to describe technical information useful to enhance product features, thus food and beverages companies take advantage of Quantitative Descriptive Analysis (QDA) [2] [3].

Since taste is not objective, compared to the sight for example, but partially subjective and affecting by the mood of the taster, objective analytical techniques have been used to support or, in some cases, replace the classical QDA, such as Electronic nose and electronic tongue [4].

Therefore the availability of a number of instrumental techniques has opened up the possibility to calibrate the sensory perception. Analytical techniques like mass spectrometry (MS) [5] [6] and gas chromatography (GC) [7] have been used, but very specific sample preparations are required (e.g. sample derivatization or volatilization).

In this study (*Paper I*) we tested the potentiality of Nuclear Magnetic Resonance spectroscopy as “magnetic tongue” to measure sensory descriptors of canned tomato samples, since NMR spectroscopy is one of the most fast, accurate and not expensive analytical technique.

In particular, we used a metabolomic NMR *fingerprinting* approach in order to differentiate the samples on the basis of their chemical composition and then correlate the fingerprints to the sensory descriptors. Final results will demonstrate that NMR spectroscopy might be a very useful tool for the characterization of sensory features of tomatoes.

In this section the project workflow and final results are reported.

3.1.1 Project workflow

The metabolomic fingerprinting approach of this project involves several steps in order to find the answer to our specific research question.

The workflow followed for this project is summarized in Figure 1.

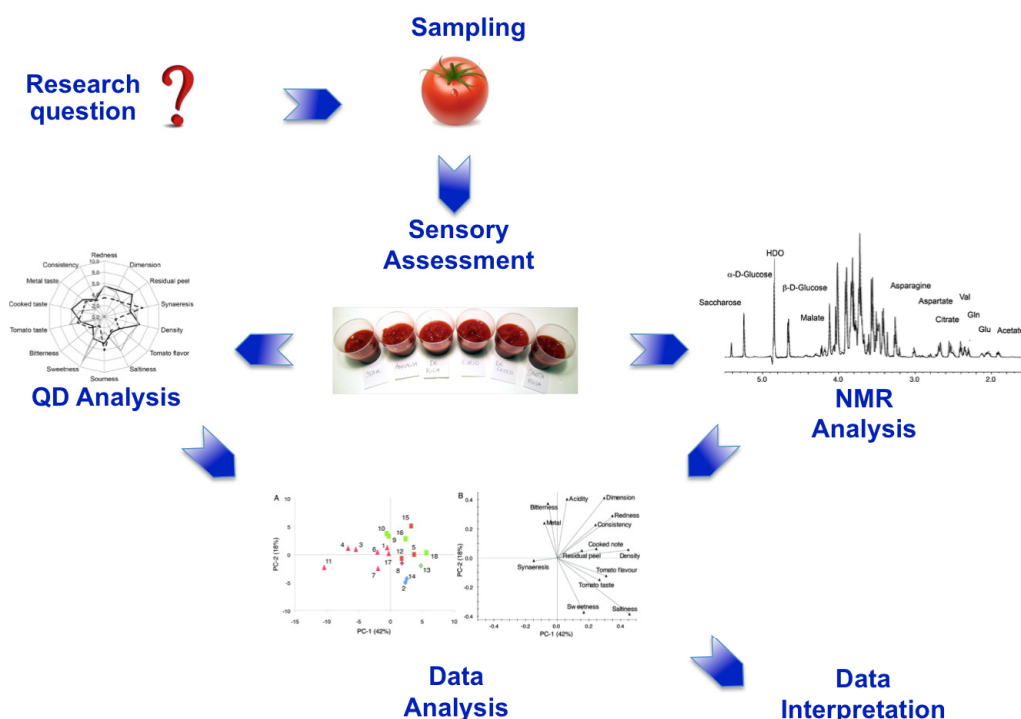


Fig.1 Illustration of Project workflow.

This study started as a “pilot” project with the aim to explore the analytical potentiality of the NMR spectroscopy as “magnetic tongue” in the analysis of eighteen canned tomato products of different brands purchased in different markets in Napoli (Italy). Sensory assessments were carried out by a panel of trained assessors, who developed a specific profile protocol for QDA containing 14 descriptors. ¹H NMR analysis was performed on the same

samples and the resulting data matrixes were analyzed using chemometric tools in order to identify the chemical signatures of sensory descriptors and predict the descriptors independently from quantitative descriptive analysis.

3.1.2 Methods and results discussion

Quantitative descriptive method

The QDA method aims to define a product's sensory profile, describing products in terms of sensory features as perceivable through five senses. In QDA descriptors were evaluated on a continuous, unlabeled, 0-10 intensity scale and then turned into numeric variables (a number between 0 and 10). Three replicates per sample were performed to minimize random errors and in particular, each subsequent replicate after one week from the previous one.

During each session a maximum of three samples were presented, according a balanced rotation plan. Then, the same samples were subjected to ^1H -NMR analysis, using parameters already described in the previous chapter.

Hierarchical Cluster Analysis

A Hierarchical cluster analysis (HCA) was performed on mean QDA parameters and on NMR data to identify similar products sharing similar sensory properties and to eventually find correlations among the different data set (Fig.2).

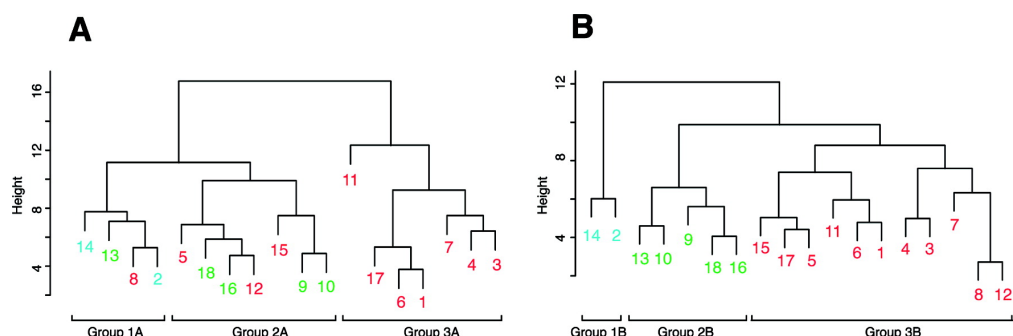


Fig.2 Dendrograms showing the similarities between products based on (A) QDA and (B) NMR. Products falling within the same group in the NMR classification are indicated with the same color.

The dendrograms describing the sensory analysis were based on unscaled sensory data, while the NMR-based dendrograms were based on PLS-DA scores of VAST scaled [8] NMR data calculated using Simca-P 11.5 (Umetrics, Umea, Sweden) as input.

Despite the fact that, the two HCAs refer to data collected by very different analytical techniques, it can be seen that there is a good global agreement between the different measurements: products of group 1B (2 and 14) are also present in group 1A, all products except one in group 2B are also present in group 2A, and all products in group 3A are also present in group 3B.

Principal Component Analysis

An exploratory and unsupervised analysis was conducted performing PCA models in parallel on unscaled sensory data and on VAST-scaled NMR data.

The PCA resulting plot has shown the QDA ability to define a sensory map of products positioning, within an overall picture, products in terms of their sensory properties.

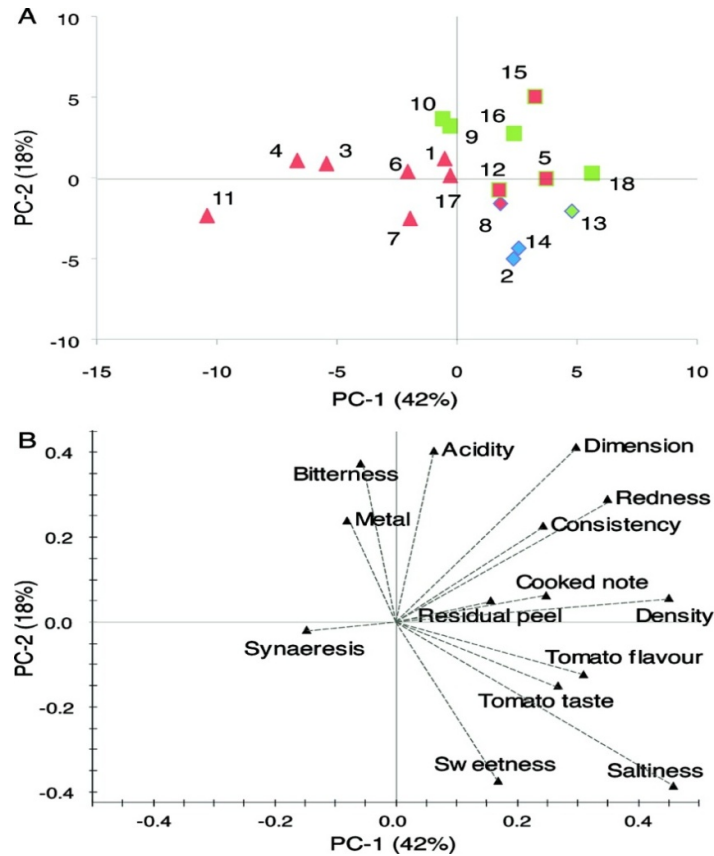


Fig.3 Score (A) and loading (B) plots of the PCA performed on sensory data. Products are colored according to NMR HCA analysis in Figure 2B. Note that none of the sensory descriptors are well described by this PCA model ($|R| > 0.5$ for all descriptors).

In particular, the scores (Fig.3A) allowed us to identify the most important sensory descriptors for products differentiation. This analysis indicates that the groups identified by the HCA share the same features and that there is no strong separation between the different groups identified.

Interpreting the loading plot (Fig.3B), it's clear that products belonging to group 1A are characterized by sweetness, by tomato taste and saltiness and by tomato flavor. Group 2A is instead characterized by a more marked redness and sourness. On the other hand, group 3A is characterized by bitterness and metal taste, having a light redness.

PCA has also been performed on the NMR data set (Fig.4), providing information about samples distribution based on their sensory descriptors.

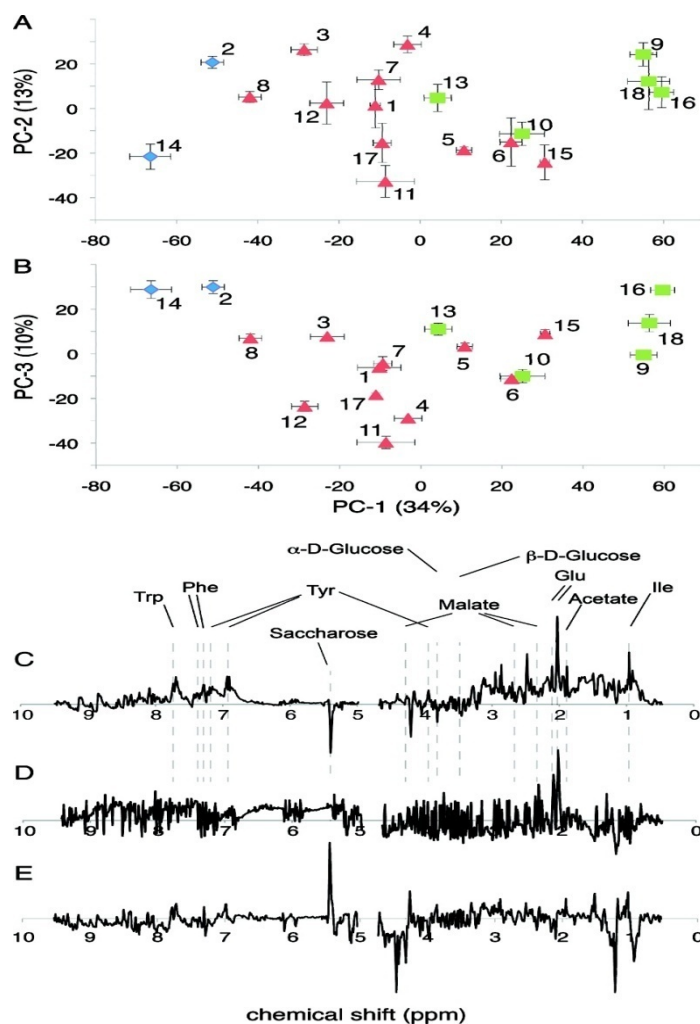


Fig.4. Score (A and B) and loading (C-E) plots of the PCA performed on NMR data. Panels A and B show the PC1_PC2 and PC1_PC3 score plots, and panels C-E show the PC1_PC3 loadings. Products are colored according to NMR HCA analysis in Figure 1B. Error bars correspond to one SE ($SE = SD/N_{1/2}$).

Interestingly the general distribution of the products in the score plots (Fig. 4A,B) in a way recall the one observed in the sensory data set (Fig. 3A).

Moreover, PC1, PC2, and PC3 vary significantly between the different products. As judged from the loading plots (Fig. 4C-E), the first PC describes the distribution of the samples based on their sweetness. In fact, negative values can be observed for signals belonging to sugars like saccharose and α - and β -D-glucose. At the same time, positive correlations can be observed for signals belonging to bitter amino acids like tyrosine, phenylalanine, tryptophane, and isoleucine. The noisy look of the second PC describes the formation of sharper NMR signals due to a decrease in viscosity. The third PC seems instead related to an increase of saccharose, isoleucine, and acetate and a decrease of tyrosine, α -D-glucose, malate, and glutamate.

The similarities in the structures of the sensory and NMR data were very encouraging and allowed us to carry on with the crucial step of the study.

In order to answer to our research question about how well NMR data can predict the sensory descriptors, we performed *Orthogonal projection to latent structures* (OPLS) [9] models using Simca-P 12.0 (Umetrics). OPLS was carried out using each sensory descriptor as the y-variable and the data were scaled to obtain unit variance and then centered.

Using this protocol, we were able to get good predictions [$Q^2(\text{cum}) > 0.5$] for bitterness, redness, density, and metal and tomato taste.

Finally, to determine the chemical components responsible for a given sensory descriptor, we have looked for all possible correlations between the NMR signals and the analyzed sensory descriptors using OPLS models.

NMR signals that showed a strong correlation ($R^2 > 0.5$) with the OPLS predictive scores for the sensory descriptors were considered as markers for the sensory descriptors.

Using this approach was possible to correlate the NMR fingerprinting of canned tomato samples to relative sensory descriptors.

For instance, sweet perception was positively correlated with saccharose (5.41 ppm) in spite of its low concentration, whereas it was negatively correlated with tyrosine (H- α 3.94 ppm), which is a known bitter amino acid.

It was interesting to note that the sensation of sweetness cannot solely be explained by the sugar content, accordingly to other data in literature [10] that reveal a strong relationship between the sensation the sensation of sweetness and the glucose/citric acid interaction.

Very surprisingly, redness was positively correlated with the presence of tryptophan (H4, H5, and H6 at 7.75, 7.17, and 7.29 ppm, respectively) and tyrosine (H3/H5 at 6.90 and H- β s at 3.06 and 3.18 ppm).

Finally, a number of signals in the region between 4.30 and 4.60 ppm and at 4.03 ppm display negative correlations with density. Unfortunately we were not able to unambiguously assign these signals, even if their chemical shifts strongly suggest that they could be attributed to sugars.

3.2 Application of “magnetic tongue” to the sensory evaluation of extra virgin olive oil

The results obtained in the previous study suggested that the metabolomic NMR fingerprinting represents a very useful tool to explore and predict sensory features of food samples. To evaluate the applicability of this methodology, the same strategies and study design were applied to predict and to measure sensory descriptors in extra-virgin olive oil, with particular attention to the quantitative measure of minor compounds related to the sensory description (*Paper II*).

Extra-virgin olive oil (EVOO) has received increasing attention over the world for their unique nutritional and healthy properties and extraordinary flavor and taste.

In the last decade, NMR spectroscopy was used to analyze extra virgin olive oil in several applications. In 2001 Mannina L. and coworkers evaluated the potential contribution of this technique to the geographical characterization of olive oils [10]. Furthermore, other analytical techniques, such as the electronic nose and the electronic tongue, in combination with multivariate analysis, have been used to verify the geographical origin and the uniqueness of specific extra virgin olive oils [11]. The recent development of NMR spectrometers (high field, cold-probe) and their performance in term of both resolution and sensitivity open new perspectives in the application of this powerful analytical technique in the analysis of extra virgin olive oil, especially in combination with multivariate data analysis.

In this section an overview of all results is reported, considering that was applied the same approach of the previous study.

The EVOO analyzed products were given from different companies in Campania region (Italy) within the EXTRABIO 2008, which is a quality prize organized by the Chamber of Commerce of Naples for olive oils from organic agriculture.

3.2.1. Methods and results discussion

Sensory profiles of the 18 samples were determined by the olive oil sensory panel of the “Laboratorio Chimico Merceologico” of the Chamber of Commerce of Naples (Italy) and eleven descriptors have been defined: fruity, leaf, grassy, bitter, pungent, sweet, almond, artichoke, apple, tomato and rosemary tastes. Then samples were prepared for ^1H -NMR analysis following The procedure reported by Segre and Mannina [12].

Hierarchical Cluster Analysis

A Hierarchical cluster analysis (HCA) was performed on mean QDA parameters and on NMR data, using Ward clustering method [13] in order to group products sharing similar sensory features (Fig.5).

Analogously to the HCA performed on sensory data, the HCA analysis performed on the NMR data revealed three main groups (Fig. 5B): there is an excellent global agreement between the different measurements but only sample 5, 11 and 18 are not grouped in the same way using QDA and NMR data.

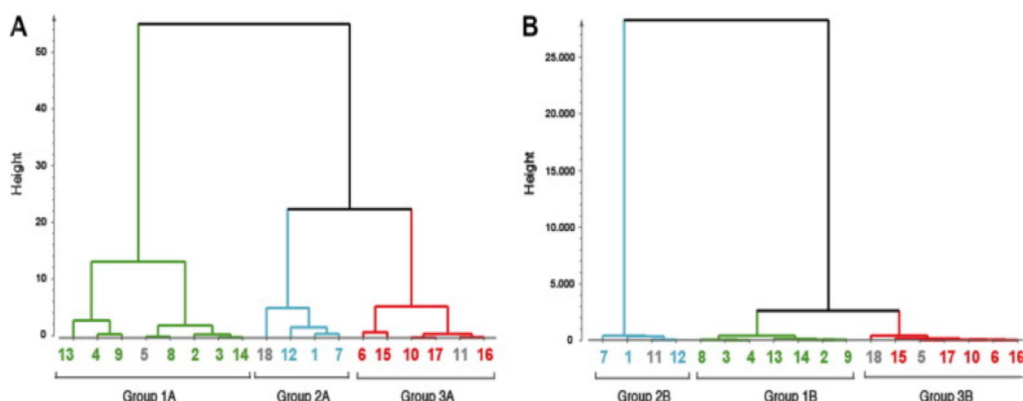


Fig.5. Dendrograms showing similarities between products based on QDA (A) and NMR (B). The dendrograms were based on the two first PCs after PCA of QDA (A) and NMR (B) data, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Principal Component Analysis

PCA models were performed on sensory data and also on the NMR data, using Simca-P 13.0 software (Umetrics, Umea, Sweden) (Fig.6). The sensory data were autoscaled while NMR data were pareto-scaled, since these data pre-treatments seems to perform better with regard to expectations.

Two principal components accounting for 65% of the variation were identified. The plot of their scores (Fig. 6A) shows the positioning of the products according to their sensory attributes and allowed the identification of the most important sensory descriptors for products differentiation.

Interestingly, a number of descriptors shows a significant correlation ($Q^2(\text{cum}) > 0.4$) with the model, indicating a high level of correlation between the sensory descriptors (Table S2 in supporting material).

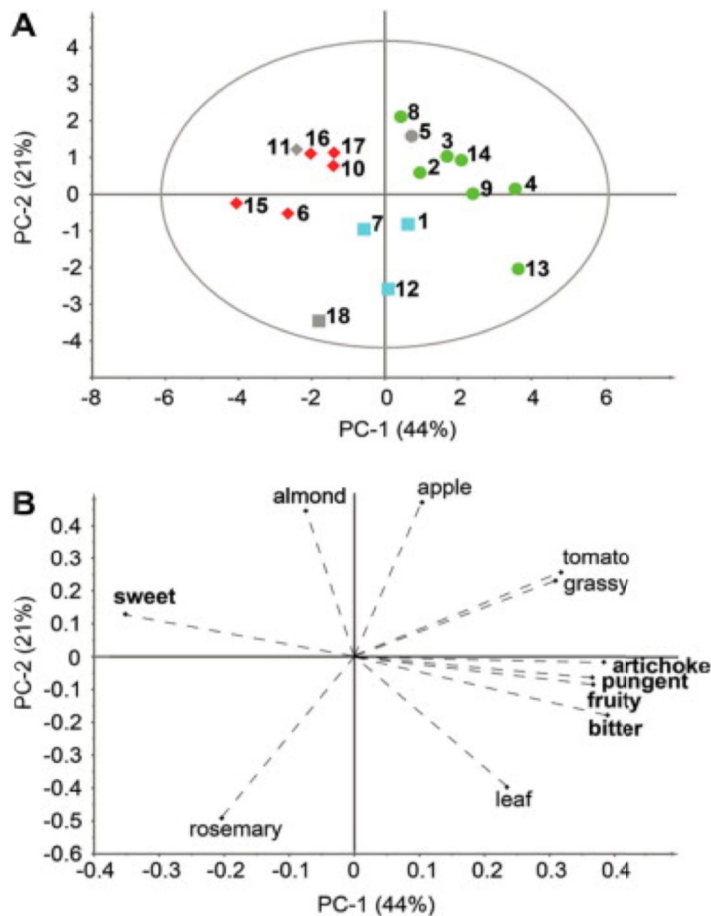


Fig.6. Score (A) and loading (B) plots of the PCA performed on sensory data. Products are coloured according to QDA HCA analysis in Fig. 5.

According to the loading plot (Fig. 6B), the transition from the left to the right of the map shows the simultaneous decrease of the sweet taste and increase of the bitter, pungent, fruity and artichoke tastes. Moreover, this plot suggests a covariance between different tastes: a higher sweetness is generally associated with a lower bitterness.

Regarding the PCA performed on the NMR data, two PCs were identified, explaining the 83.6% of the variation, The general distribution of the products in the score plots (Fig. 7A) in a way recall the one observed in the sensory dataset (Fig. 6A).

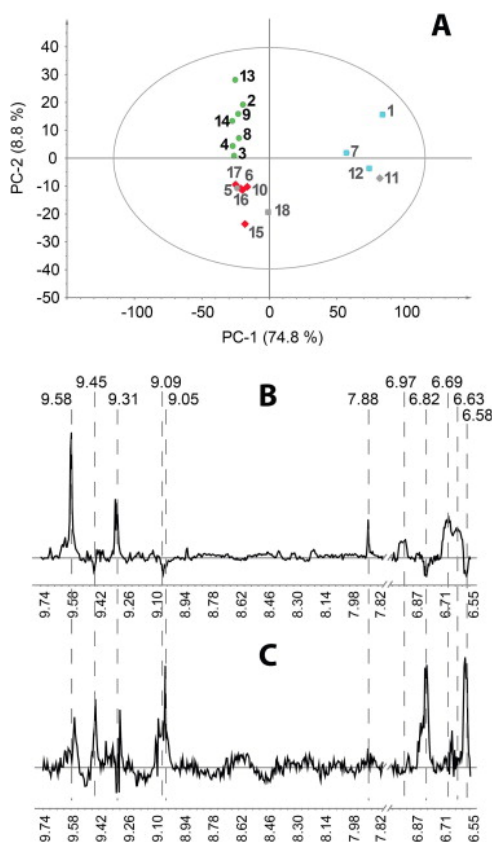


Fig.7. Score (A) and loading (B and C) plots of the PCA performed on NMR data. (B and C) show the PC1 and PC2 loadings. Products are coloured according to Fig. 5.

For example, Group1B and Group 3B are mapped opposite to each other, suggesting that PC2 in the PCA computed with the NMR data is in a way related to the sweetness/bitterness of the samples. On the other hand group 2B is very well separated from the other two groups along PC1, suggesting that this PC is related to apple, rosemary and leaf tastes.

Samples contained in this group are characterized by higher concentration of molecules having signals at δ_H 9.58, 9.31, 7.88, 6.97, 6.69 and 6.63 ppm, and low concentration of molecules having signals at δ_H 9.45, 9.09, 9.05, 6.82 and 6.58 ppm, relative to Groups 1B and 3B.

Looking at the loading plot of second principal component (Fig. 7C), it is clear that Group 1B contains higher concentrations of molecules having signals at δ_H 9.45 (the aldehyde proton of the trans-2-hexenal), 9.09, 9.05, 6.82, 6.58, whereas lower concentrations of those molecules are present in the Group 3B.

Signals around δ_H 9.10 could also be tentatively assigned to protons of the dialdehyde form of secoiridoids, and signals at δ_H 6.5–6.8 to phenyl alcohols moieties (tyrosol and hydroxytyrosol) of oleuropein and ligstroside aglycons [14]. Unfortunately other signals were not unambiguously assigned.

In order to determine which chemical components are responsible for a given sensory descriptor, we have looked for all possible correlations between the NMR signals (in the considered spectral regions) and the analyzed sensory descriptors using OPLS models.

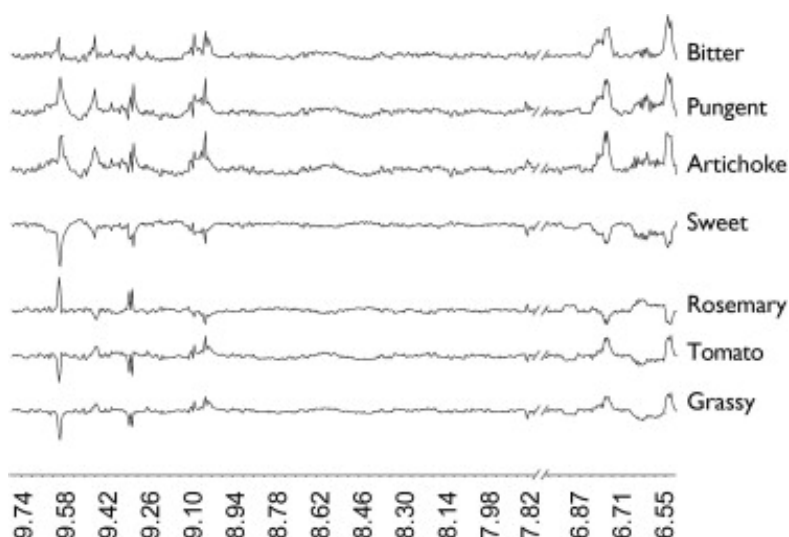


Fig.8 Loading plot of the predictive component of OPLS models with $Q^2 > 0.4$.

If we have a look at the loading plots of the predictive component of each OPLS model with $Q^2 > 0.4$ (Fig.8), bitter, pungent and artichoke tastes are highly correlated ($R > 0.63$) and display a very similar profile. They all show a strong anti-correlation to sweet taste ($R < -0.6$), while tomato and rosemary tastes display inverse profiles relative to each other.

Inspecting in depth the OPLS data, especially regarding the less intense signals, we noticed that the lack of hexenal seems to increase sweet, tomato, grassy and fruity tastes, whereas the increment of its concentration increases the perception of leaf and rosemary tastes. Secoiridoids (signals around δ_H 9.10), trans-alk-2-enals (δ_H 6.84) and 4-hydroxy-trans-alk-2-enal (signals at δ_H 9.58) are related to the sweet/bitter relationship of the EVOO.

Finally, in according to other papers in literature [15] the trans-2-hexenal (δ_H 9.45 ppm) correlates to fruity taste of the olive oil.

3.3 NMR profiling of cocaine seizures

Nowadays, there are hundreds of drugs that are commonly abused, and many of these are legally sanctioned by many countries.

One of the most widely illicit drug used is cocaine; it is an extract of the leaves of the coca plant, where it is found as the main natural alkaloid.

In addition to cocaine, coca paste contains also small percentages of other compounds, mostly alkaloids [16], which can be considered a “fingerprint” of the sample. The coca samples chemical “fingerprint” can depend on several factors: which kind of plant minor compounds were extracted from, where it was cultivated, and which procedures were used for extraction and purification. Thus the presence of these minor compounds in cocaine can provide information about the geographic origin, but also about the area and the period of the cocaine traffic.

A recent search of the books for sale on the topic of drug and pharmaceutical analysis at *Amazon.com* had more than 300 entries. Indeed forensic drug analyst and toxicologists utilize methods and techniques that continue to develop rapidly. In particular, the GC–MS technique currently dominates forensic analysis [17]. Since NMR represents a unique methodology for performing chemical identification and quantification, that does not require specific sample preparation or the use of other analytical technique (e.g. chromatographic analysis), here a strategy based on ^1H NMR spectral analysis in conjunction with multivariate analysis is presented. This project aims to identify the chemical “fingerprint” of cocaine samples seized at different times and in different places in Naples. Thus, using NMR fingerprinting approach we could be able to improve investigation strategies.

3.3.1. Methods and results discussion

In this study we performed an NMR analysis of 54 chlorohydrate cocaine samples seized by the Police Department of Naples in different areas of Naples, during the year 2006.

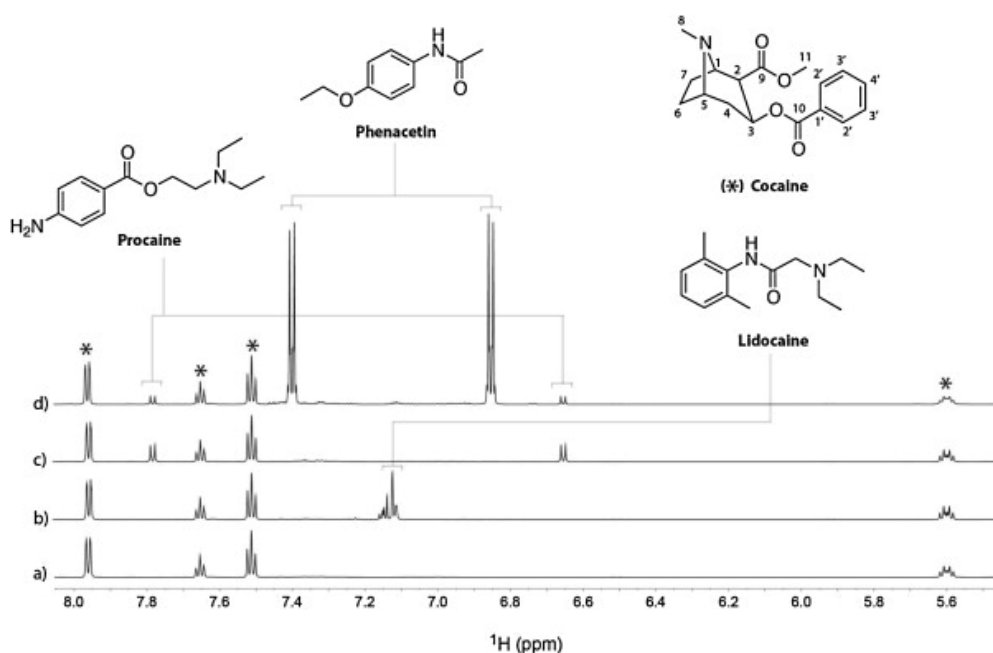


Fig.9 Comparison of an expanded region of four samples of seized cocaine. (a) Example of a sample of pure cocaine. Asterisks indicate cocaine signals. (b) Sample having lidocaine as cutting agent. (c) Sample having procaine as cutting agent. (d) Sample having procaine and phenacetin as cutting agents.

In the spectra the main signals belong to the cocaine hydrogens (complete assignment is reported in the table S1 in Supplemental Material), but other signals are observed, such as cutting agents (adulterant/diluent) that could be mixed to the cocaine samples before drug dealing (Fig.9). It was possible to identify all the signals of cocaine and cutting agents with the help of literature data [11].

Moreover, cutting compounds were easily detected and quantified by a simple ^1H NMR spectrum and can be used as an investigative support to determine a common origin and a distribution channel of the illicit drug.

But the main goal of this study is to analyze the “fingerprint” of cocaine samples in order to identify the areas and period in which a given consignment of drug has been trafficked.

For this reason we operated a selection of the spectral regions including information related only to the minor compounds, without interference from cocaine and adulterant/cutting agent signals.

Statistical Total Correlation Spectroscopy (STOCSY)

STOCSY technique was used for determining metabolic connectivity between different molecules as well as for structural assignment in NMR spectra of complex mixtures [18] (see figure S1 in *Supporting material*). STOCSY was performed on NMR data set (54 ^1H NMR spectra) to generate a pseudo-2D NMR spectrum, which displays correlation among the intensities of various peaks across the whole spectrum. In order to perform this technique the spectra were divided in equally sized bins (0.005 ppm), reducing the data set number of variables and the STOCSY time for calculation.

In the STOCSY plot correlations among following agents were detected: lidocaine, phenacetin, diltiazem, sugars, procaine, MDMA (3,4-methylenedioxy- N-methylamphetamine), paracetamol and caffeine and then all the spectral regions in which those signals were present were excluded.

Thus, 7 spectral intervals that only contain signals of the minor (fingerprint) components were considered in the multivariate data analysis.

Prior data analysis, this spectral intervals of each sample were normalized with respect to the area of a reference cocaine peak, since the amount of the minor components is somehow related to the amount of cocaine. Particularly, the cocaine triplet at δ H 7.65 ppm (H40) has been used, since it is an isolated peak in all 54 samples.

The normalized data matrix was then used for hierarchical cluster analysis (HCA), which shows 5 main groups that can be considered as 5 different drug consignments (Fig.10).

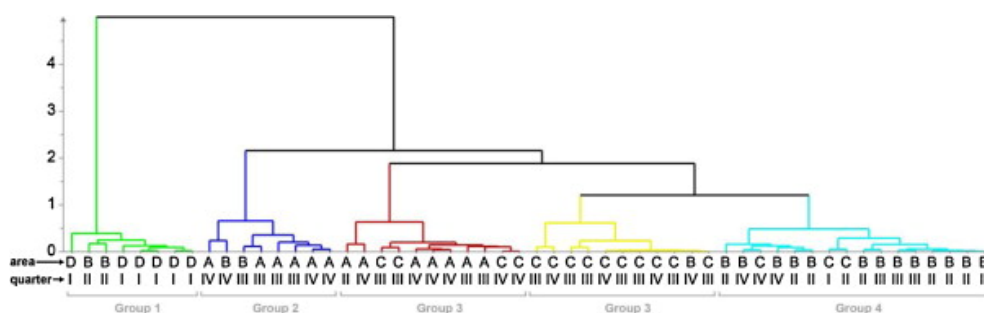


Fig.10 HCA dendrogram based on NMR data, showing similarities between samples. Samples falling within the same group are indicated with the same color. Area (A, B, C and D) and quarter of the year (I, II, III and IV) of seizure are also reported.

Principal component analysis

PCA was also performed on the same dataset (Fig. 11) in order to get insights

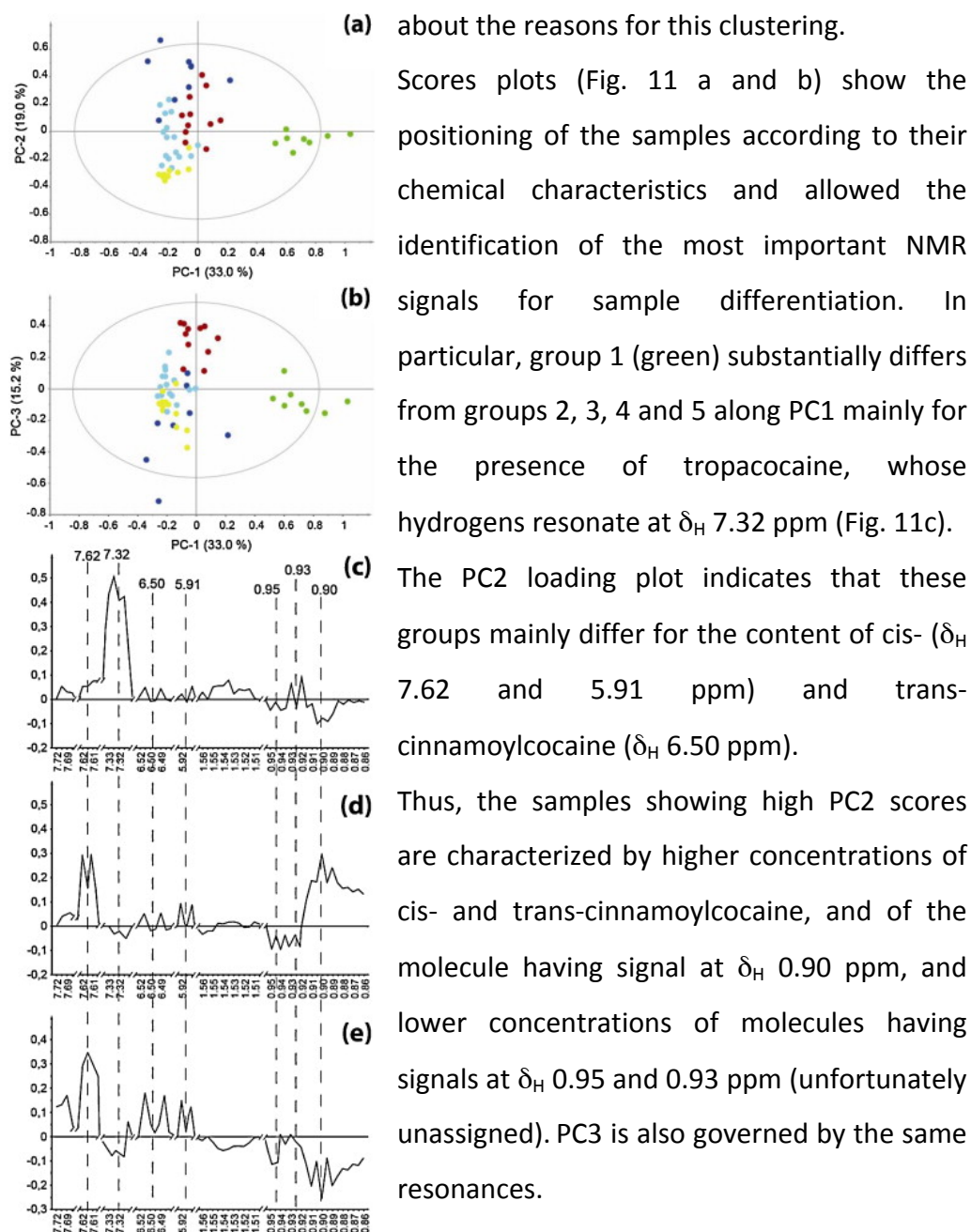


Fig.11 PCA score (panels a and b) and loading (panels c–e) plots.

Projection to Latent Structures Discriminant Analysis

PLS-DA models were performed in order to determine how well the identified consignments of drug can be predicted by NMR.

The validation of the PLS-DA models was obtained using the permutation test (Simca-P 13.0, Umea, Sweden) in which a total of 400 models were calculated using randomly permuted Y variables (Fig. S2 in *supporting material*).

The resulting Q² and R² values, describing the predictive ability and the reliability of the fitting, respectively, were plotted and compared with the Q² and R² values obtained from the real model. The substantial decrease of both parameters Q² and R² (vertical axis interception point of the Q² and R² regression line resulted both with negative values) enforced the statistical validity of the obtained PLS-DA model.

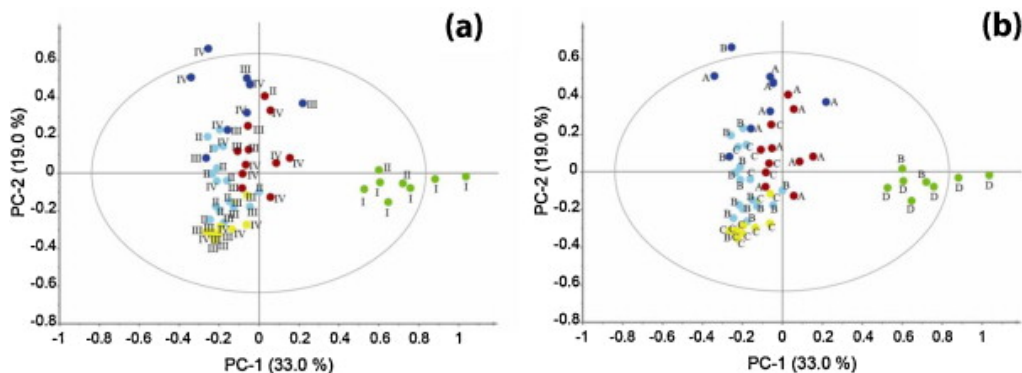


Fig.12 PCA score-plots annotated with (panel a) year quarter and (panel b) area of seizure.

In the plot on the left all samples are labeled according to the quarter of the year in which cocaine samples were seized (I, II, III and IV) (Fig. 12a), on the right according to four macro areas of seizure (Fig. 12b). Macro areas were defined based on information from Naples Police Department, which assigned

each area to a single criminal organization or clan, responsible of cocaine seizure.

Samples of group 1 (green) have been seized mainly in area D of the county. Analogously, samples of groups 2 (blue), 3 (red), 4 (yellow) and 5 (cyan) are seized mainly in areas A, A/C, C, and B, respectively.

This suggests that each clan stocks up with drug consignments independently from the others and that it actually controls a given area in an exclusive manner. It is also possible to monitor the samples that, for some reason, move from the area where most of the consignment to which they belong has been sold.

Finally, it was also possible to identify the consignments for this drug trafficked in different period. For instance, group 1 (green) have been sold mainly in the first (I) quarter, groups 2, 3 and 4 have been trafficked in the third (III) group 5 (blue) has been sold all year long.

3.4 Conclusions

The results obtained in all three case studies suggest that NMR could be a very an exploratory and predictive useful tool for the characterization of some sensory features of canned tomato and extra virgin olive oil, or generally for food samples. We were able to correlate the NMR metabolomic fingerprints recorded for canned tomato samples to the sensory descriptors bitterness, sweetness, sourness, saltiness, tomato and metal taste, redness, and density and NMR metabolomic fingerprints of extra virgin olive oil to the sensory descriptors: tomato, bitter, pungent, rosemary, artichoke, sweet, grassy and leaf.

The reported results are very promising and pave the way to a more careful analysis on a wider number of samples.

Furthermore, the potentiality of NMR spectroscopy, in combination to multivariate data analysis, has also demonstrated to improve and support police investigations. In particular, the NMR fingerprinting approach of cocaine samples analysis allowed us to identify the spectral regions that, at least in this investigation, can be used for grouping of seized cocaine samples. These regions were very useful especially in providing information about place and date of trafficking. Moreover, the statistical multivariate analysis allowed the identification of key minor components responsible of this grouping. The results indicate that this study strategy can assist tactically (evidential/judicial) and strategically (intelligence) the investigators.

3.5 REFERENCES

- [1] Roper, S.D. Cell communication in taste buds. *Cellular and Molecular Life Sciences*, **2006**, 63, 1494–1500.
- [2] Noble, A.C. Sensory methods of flavor analysis. In Food Flavour Technology; Taylor, A. J., Ed.; Sheffield Academic Press: Sheffield, United Kingdom, **2002**; pp 252_275.
- [3] Stone, H., Sidel, J.L. Quantitative descriptive analysis: Developments, applications, and the future. *Food Technol.*, **1998**, 52 (8), 48–52.
- [4] Deising, A.K., Stone, D.C., Thompson, M. Application of electronic noses and tongues in food analysis. *Int. J. Food Sci. Technol.*, **2004**, 6, 587–604.
- [5] Aishima, T. Correlating sensory attributes to gas chromatography mass spectrometry profiles and e-nose responses using partial least squares regression analysis. *J. Chromatogr., A*, **2004**, 1054, 39–46.
- [6] Zanol, M.I., Rambla, J., Chaib, J., Steppa, A., Medina, A., Granell, A., Fernie, A. R., Causse, M. Metabolic characterization of loci affecting sensory attributes in tomato allows an assessment of the influence of the levels of primary metabolites and volatile organic contents. *J. Exp. Bot.*, **2009**, 60, 2139–2154.
- [7] Taylor, A.J., Rob, S.T., Linforth, R.S.T. Direct mass spectrometry of complex volatile and non-volatile flavour mixture. *Int. J. Mass Spectrom.*, **2003**, 223_224, 179–191.
- [8] Keun, H.C., Ebbels, T.M.D., Antti, H., Bollard, M.E., Beckonert, O., Holmes, E., Lindon, J.C., Nicholson, J.K. Improved analysis of multivariate data by variable stability scaling: Application to NMR-based metabolic profiling. *Anal. Chim. Acta*, **2003**, 490, 265–276.

- [9] Trygg, J., Wold, S. Orthogonal projections to latent structures, O-PLS. *J. Chemom.*, **2002**, 16, 119–128.
- [10] Stevens, M.A., Kader, A.A., Albright-Holton, M., Algazi, M. Genotypic variation for flavor and composition in fresh market tomatoes. *J. Am. Soc. Hortic. Sci.*, **1977**, 102, 680–689.
Mannina L., Patumi M., Proietti N., Bassi D., Segre A.L. Geographical Characterization of Italian Extra Virgin Olive Oils Using High-Field ^1H NMR Spectroscopy. *J. Agric. Food Chem.*, **2001**, 49 (6), pp 2687–2696.
- [11] Geographical origin and authentication of extra virgin olive oils by an electronic nose in combination with artificial neural networks. Maria S. Cosioa M.S., Ballabio D., Benedetti S., Gigliotti C. *Analytica Chimica Acta*, **2006**, 567, 2, 202–210.
- [12] Segre A.L., Mannina L. ^1H NMR study of edible oils. *Recent Research Developments in Oil Chemistry*, **1997**, 1, 297–308.
- [13] Ward J.H.Jr. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, **1963**, 48, 236–244.
- [14] Christophoridou S., Dais, P. Detection and quantification of phenol compounds in olive oils by high resolution ^1H nuclear magnetic resonance spectroscopy. *Analytica Chimica Acta*, **2009**, 633, 283–292.
- [15] Morales M.T., Alonso M.V., Rios J.J., Aparicio, R. Virgin olive oil aroma: Relationship between volatile compounds and sensory attributes by chemometrics. *Journal of Agriculture and Food Chemistry*, **1995**, 43, 2925–2931.
- [16] Isenschmid D.S. Cocaine, in: Wiley Encyclopedia of Forensic Science, John Wiley and Sons, New York, **2009**.

- [17] Moore J.M., Casale J.F. In-depth chromatographic analyses of illicit cocaine and its precursor, coca leaves, *J. Chromatogr. A*, **1994**, 674, 165–205.
- [18] Cloarec O., Dumas M.E., Craig A., Barton R.H., Trygg J., Hudson J., Blancher C., Gauguier D., Lindon J.C., Holmes E., Nicholson J. Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic ^1H NMR Data Sets, *Anal. Chem.*, **2005**, 77, 1282–1289.

4 NMR METABOLOMICS OF HUMAN CANCER CELL LINES

This section describes the preliminary results, not yet published, of a study carried out at the University of Copenhagen, Department of Food Science, Faculty of Life Science, under the supervision of Assoc. Prof. Francesco Savorani and Prof. Søren Balling Engelsen.

This study deals with the analysis of the metabolomic profiles of human cancer cell lines, using NMR spectroscopy combined with chemometrics tools. The aim of the study was to develop an experimental protocol for an efficient harvesting, quenching and extraction of cellular metabolites of the HTC-116 human adherent cancer cells. The analysis of the metabolome has been focused to understand the *in vitro* actions of novel anticancer drugs.

The experimental protocol was developed in collaboration with Dr. Pasquale Zizza and Dr. Annamaria Biroccio (Experimental Chemotherapy Laboratory, Regina Elena Cancer Institute, Rome, Italy).

4.1 Introduction

In the last decade, *metabolomics* studies have been performed on different biofluids (e.g. plasma, serum, urine and cerebrospinal fluid) with successful results, showing applications in many areas, such as biomarker discovery, clinical studies, drug efficacy and toxicity evaluations, disease diagnosis [1][2][3]. However, recent developments in the use of metabolomics involve the characterization and interpretation of the cell metabolome, starting from prokaryotes (especially *Escherichia Coli*) to eukaryotes cell lines (yeast or mammalian cells) [4][5].

Complementary to the classic biofluid analyses, for *in vivo* models the metabolomic profiles of cells represent a very powerful tool to understand how the local metabolism and biochemical pathways could be influenced by pathologies or external or internal stimuli. For the more simple *in vitro* models, such as cell culture, the analysis of the cellular metabolome could also provide important information for the development of models of biological pathways and networks. *In vitro* cell metabolomics analysis offers several advantages: experimental variables are easier to control, greater reproducibility, less expensive and easier to interpret than analysis of animal models and human subjects [6].

The use of mammalian cells is emerging in the metabolomics field and it is worthwhile to understand the molecular mechanism of disease progression, the cellular response to drug treatments [7] and the cell culture monitoring [8].

In particular, the identification and characterization of cancer cell metabolomic signatures could play an important role in the early diagnosis as well as in following therapeutic response, making possible to map the drug action into metabolic pathways [9].

Colon carcinoma is the third most commonly diagnosed cancer in the world and the second most common cause of death from cancer [10]. The analysis of metabolic profiles of human colon cancer cell lines – using NMR spectroscopy-based metabolomics – aims to provide a comprehensive assessment of the alterations in the metabolite levels in cells and could produce important information on *in vitro* actions of drugs pointing in their rapid incorporation into novel therapeutic settings.

For a holistic understanding of cancer cells metabolome, the choice of the right protocol is crucial for efficient harvesting, quenching and extraction of cellular metabolites. This study reports the development of an optimized and standardized protocol for NMR metabolomics study of HTC 116 human colon cancer cell lines.

4.2 Methodology of sample preparation and extraction

Cell metabolomic experiments can be divided into different step: cell growth, quenching and metabolites extraction.

To prepare cells for metabolomics investigations, quenching aims to inactivate intracellular enzymes and arrest metabolic activity as rapidly as possible to avoid metabolite degradation and alteration of the sample composition, since a number of metabolic reactions occur in seconds. This method stops metabolism by rapid deep cooling of the cells; the lower the temperature the slower the turnover rate of all the enzymes within the cell and hence the more efficient the quenching process. A variety of studies involving low temperature quenching are reported in literature. Cell quenching process of mammalian cells can be achieved in cold-methanol on its own [11] or buffered with reagents like ammonium bicarbonate (AMBIC) as described by Sellick et al. [12].

The choice of the method for quenching depends from the cellular composition (cell membrane and cell wall structure) and cell size also may influence the efficiency of quenching and the rate of metabolite leakage. Indeed, methanol-based solutions damage the cell membrane and induce leakage of intracellular metabolites from bacteria and yeast [13] [14].

This effect is more marked in mammalian cells, since they are more fragile than bacteria or yeast due to the lack of a cell wall. In contrast, Dietmair et al. proposed quenching with cold NaCl; this method did not damage cells and effectively halted conversion of ATP to ADP and AMP, indicative of metabolic arrest [15]. However, in the case of adherent cells, liquid nitrogen freezing is considered an optimal way to stop enzymatic activity, seeking to avoid the leakage of intracellular metabolites and to maintain cell integrity [16].

The less reproducible step among sample preparation procedures is intracellular metabolites extraction. According to the holistic metabolomic view, the extraction procedure must be adequate in order to ensure the simultaneous detection of a large number of metabolites. Generally polar organic solvents extract polar compounds, whereas non-polar solvents such as chloroform or dichloromethane allow the extraction of lipids and other hydrophobic compounds. Sometimes, a mixture of polar and non-polar solvents conveniently allows the extraction of both classes of metabolites. Several different extraction methods have been reported in literature including cold solvents (50-60% methanol, 100% methanol, methanol/chloroform), hot solvents (methanol, ethanol) and solvents with extreme pH values (KOH and perchloric acid) [17] [18] [19].

For the complete analysis of a cell culture, it is important to measure both extracellular (*footprint*) and intracellular (*fingerprint*) metabolic profiles. **Metabolic footprinting** (*exo-metabolome analysis*) is technically simple because it only requires centrifugation to separate culture media and cells before the analysis. **Metabolic fingerprinting** (*endo-metabolome analysis*), although much more technically challenging because it requires metabolite extraction from cells, provides more complete information about cellular

metabolic processes [5].

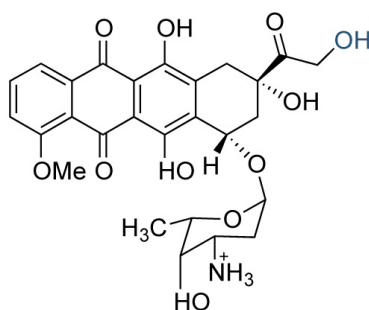
In this study a non-selective, non-destructive and efficient extraction procedure has been developed considering the metabolomics study objective in an untargeted NMR analysis. NMR spectroscopy permits the detection of hydrophilic and lipophilic intracellular metabolites with an adequate sample preparation. On the other hand, not all procedures are suitable for NMR metabolomics study. Therefore, the use of aqueous mixtures with methanol and chloroform as extraction solvents, has allowed detecting as many metabolites as possible by NMR analysis.

The aim of this study was to develop a sample preparation protocol, including cell growth, separation of cells from medium, quenching and intracellular metabolites extraction for HTC-116 human adherent cancer cells. Different amount of cells, metabolites quenching with liquid N₂ and cold methanol, Methanol/chloroform/H₂O metabolites extraction method were tested. The metabolites' leakages induced by quenching procedures were compared and the extraction efficacy was evaluated using Nuclear Magnetic Resonance spectroscopy. The developed experimental protocol was used to perform colon cancer cell metabolic profiling and to understand *in vitro* actions of novel anticancer drugs.

4.3 Anticancer drugs and DNA G-quadruplex binders

Three different drugs were tested on HTC-116 colon cancer cell lines: the Adriamycin, a popular antineoplastic agent and commonly used in chemotherapy, RHPS4 and *Compound 3*, novel anticancer drugs which are DNA quadruplex ligands, which can induce tumor-cell death.

The anthracycline antibiotic doxorubicin (trade name: Adriamycin) (Fig.1), originally isolated from the fungus *Streptomyces peucetius* [20] is a chemotherapeutic agent with strong activity against a wide range of human malignant neoplasms including acute leukemia, non-Hodgkin lymphomas, breast cancer, Hodgkin's disease and sarcomas [21].



doxorubicin (adriamycin)

Fig.1 Chemical Structures of Adriamycin

Adriamycin acts through DNA *duplex* intercalation (Fig.2) and inhibition of macromolecular biosynthesis [22][23]. Furthermore, it inhibits the progression of the enzyme topoisomerase II, which relaxes supercoils in DNA for transcription [24]. Apart from side-effects that are common to many cancer chemotherapeutics, i.e. hematopoietic suppression, nausea and vomiting and alopecia, the clinical usefulness of doxorubicin is limited largely by a cumulative dose-related cardiomyopathy that manifests itself as congestive heart failure [21].

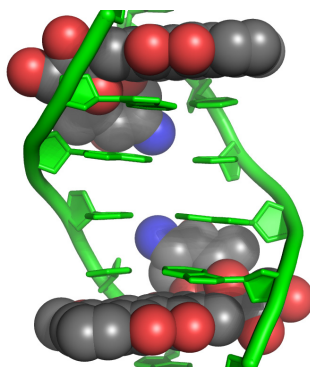


Fig.2 Cartoon diagram of two doxorubicin molecules intercalating the DNA *duplex*.

In the last decade, targeting of DNA secondary structures, such as *G-quadruplexes*, has been considered an appealing opportunity for drug intervention in anticancer therapy [25]. Recent findings have unambiguously demonstrated that DNA G-rich sequences can adopt a *G-quadruplex* folding in living cells, thus further validating them as crucial targets for the treatment of human disorders such as cancers [26].

G-quadruplex DNA (G4-DNA) structures are four-stranded helical DNA (or RNA) structures, comprising stacks of G-tetrads, which are the outcome of planar association of four guanines in a cyclic *Hoogsteen* hydrogen-bonding arrangement. From the biological point of view, G4-DNAs are widespread in the genome and it seems they play a role in a number of processes, such as replication, recombination transcription and translation [27] and are also found in telomeric regions [28].

Telomeric DNA consists of tandem repeats of simple short sequences, rich in guanine residues, which can form *G-quadruplexes*. Telomeres protect the ends of the chromosome from damage and recombination, and their shortening is implicated in cellular senescence.

The elongation of telomeric DNA operated by the enzyme telomerase leads cancer cells to an infinite lifetime.

The inhibition of telomerase, which is over-expressed in about 85% of tumors, represents the forefront of research for new effective anticancer drugs. Since this enzyme requires a single stranded telomeric primer, the formation of G-*quadruplex* complexes by telomeric DNA inhibits the telomerase activity.

Thus, cells stop replicating and enter a senescence phase which precedes apoptosis.

In this respect, it has been found that small molecules that stabilize G-*quadruplex* structures are effective telomerase inhibitors and can be considered as novel drugs in anticancer therapy [29].

In order to discover G4 binders displaying good drug-like profiles and the ability to induce a DNA-damage response at telomeres of cancer cells, a receptor-based virtual screening (VS) campaign approach was carried out, using the human telomeric G-quadruplex (Tel24) as target.

As result of this inspection, *Compound 3* (Fig.3) displayed impressive G4 binding and stabilizing properties.

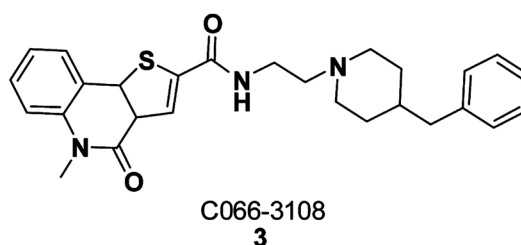


Fig.3 Chemical Structures of the *Compound 3*, newly identified Human Telomeric G4 Binding Agents

Furthermore, specific biological assays (Fig.4) showed that *Compound 3* is potent in inducing selective DNA-damage at telomeres of cancer cells and not in normal ones [30].

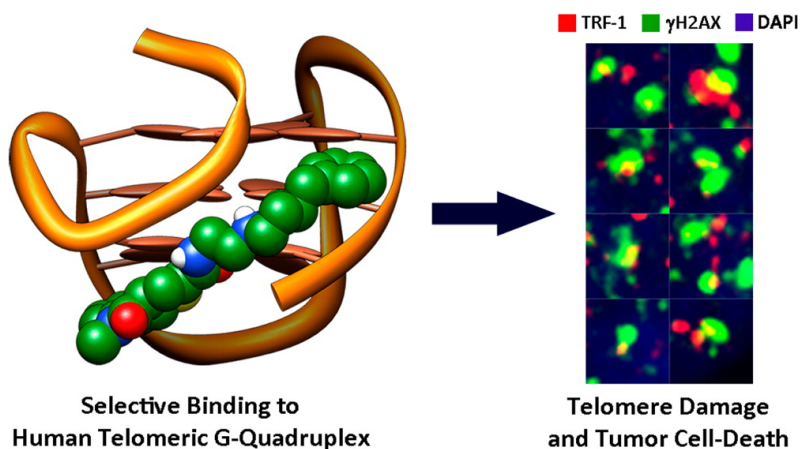


Fig.4 Groove binding mode of the *Compound3* to Tel24. On the right an enlarged view of telomere-dysfunction induced foci (TIFs) concluding that the tested compound caused telomere damage.

A challenge for the development of *G-quadruplex* interacting molecules is the relatively poor selectivity for binding to quadruplex versus duplex DNA, causing acute cell kill. The pentacyclic acridine, RHPS4 (Fig.5) has a high selectivity for quadruplex DNA structure [31].

RHPS4 inhibits telomerase at sub-micromolar levels and exhibits a wide differential between telomerase inhibition and acute cytotoxicity [32].

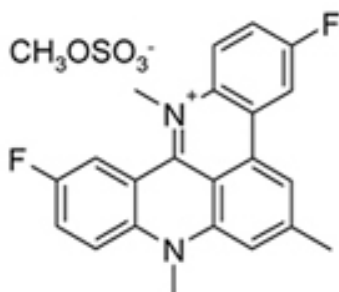


Fig.5 Chemical Structures of the pentacyclic acridine, RHPS4.

In the last decade several studies have been conducted on the fluorinated polycyclic methylacridinium salt, which describe the interaction and stabilization of quadruplex DNA and show RHPS4 enhanced binding to higher ordered DNA structures (triplex/quadruplex) over duplex and single-stranded DNA [33][34][35].

Therefore, in this study Adriamycin, as well known DNA *duplex* anticancer drug, was tested on HTC-116 human colon cancer cell lines and compared with RHPS4 and *Compound 3*, as novel human telomeric G4 **selective** binding agents.

These two anticancer drugs interact both with the same target, but RHPS4 acts by telomeric DNA G-quadruplex end-stacking (Fig.6), while *Compound 3* by groove binding [30].

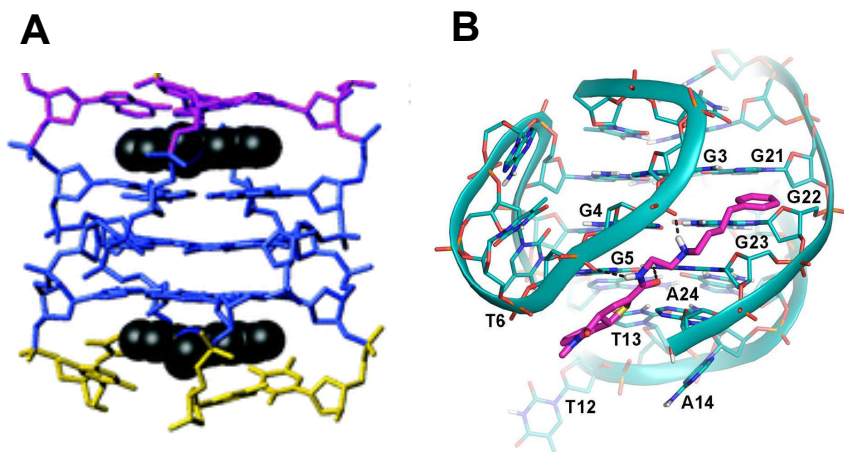


Fig.6 Binding mode of RHPS4 (A) and of Compound 3 (B).

4.4 Experimental design

Two groups of samples were analyzed in this study (Fig.7):

- **Control group** – HTC-116 human colon cancer lines not treated with anticancer drugs (10 samples).
- **Treatment group** – the same cell lines treated with three different anticancer drugs (6 Adriamycin samples, 6 *Compound 3* and 6 RHPS4 samples). The most efficient dose and drug exposure duration time of cell culture, showing therapeutic effect were chosen for each drug treatment.

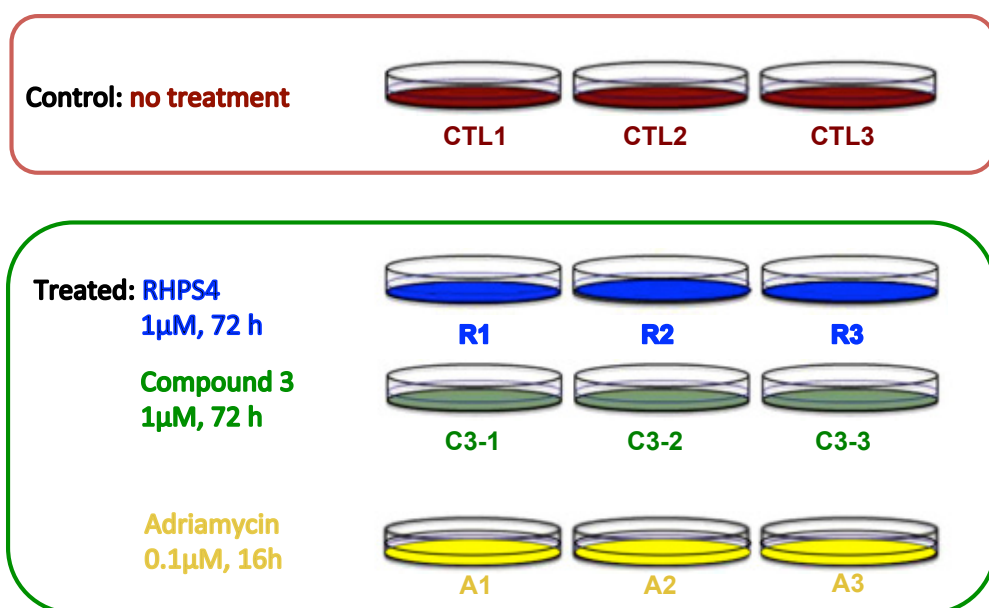


Fig.7 Overview of the experimental study.

Since it is important to ensure reproducibility when carrying out metabolomics experiments to glean biologically meaningful information, all experiments were carried out in triplicates to test the experimental variability. Three batches of sample growth and treatment were collected to perform a statistically significant analysis, including inter-batch control samples, in order to monitor the variability among the batches (Fig.8).

For the *endo-metabolome* analysis, metabolites extraction from cells was carried out. Thus, hydrophilic and lipophilic metabolites profiles were analyzed for each sample (28 samples for the hydrophilic extract, 28 samples for the lipophilic extract)

For the *exo-metabolome* analysis, growth media samples, separated from the cell culture by centrifugation, were analyzed (28 samples in total) in order to study the result of an interchange of metabolites between cells and the culture medium (uptake of substrates, excretion of metabolic products).

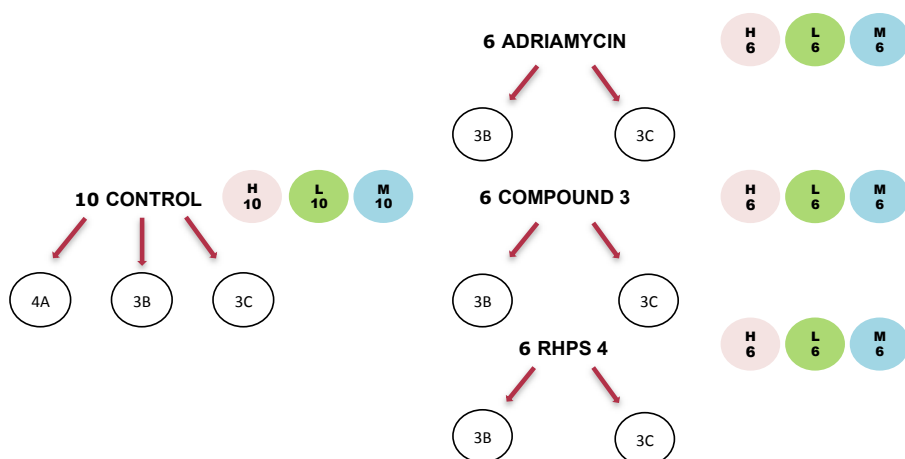


Fig.8 Scheme of the experimental study with samples size details. For each sample hydrophilic extract (H), lipophilic extract (L), cellular growth medium (M) were collected. Different batches of samples are indicated with letters A (first batch), B (second batch); C (third batch).

4.5 Materials and methods

Chemicals

Dulbecco's Modified Eagle's medium, high glucose (DMEM/HIGH Glucose) with L-Glutamine were purchased from Euroclone (MI, Italy), penicillin and streptomycin solution for cell culture were purchased from Gibco (NY, USA). Fetal bovine serum (FBS) was purchased from Thermo Scientific (Hyclone™). Crystal PBS BUFFER (0.01 M Phosphate buffer, 0.0027 M KCl e 0.14 M NaCl, pH 7.4 at 25 °C) was purchased from Bioline (TR, Italy).

All other reagents were of analytical grade. Deuterium oxide (D₂O, 99.8%D) and deuterated Chloroform (CDCl₃, 99.8%D) was obtained from Sigma-Aldrich (St. Louis, MO).

Cell Culture

Human colon cancer cell lines HCT116 were purchased from American Type Culture Collection (ATCC – Manassas, VA, USA). Cells were grown as monolayer cultures in High glucose (4.5 g/l) Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% FBS, L-Glutamine (2 mM), penicillin (1 U/ml) and streptomycin (1 mg/ml) at 37°C in a humidified atmosphere of 5% CO₂.

For *exo*-metabolome analysis, cells (30×10^6) were seeded in 150 mm tissue culture dishes (Fig.9). About 96h after seeding, the culture medium (15 ml) was recovered and centrifuged at 300 x g for 10 min. The resulting supernatant was aliquoted, frozen in liquid nitrogen and stored at -80°C until the analysis.

After medium removal, the cells were collected by scraping in 10 ml PBS and

used for the *endo*-metabolome analysis.

All the experiments were carried out in triplicates to test the experimental variability.

Anticancer drug treatments

RHPS4 and *Compound 3* were added to cell cultures 24 h after seeding. Cells were exposed to the drug treatment for 72h with 1 μ M final concentration.

Doxorubicin was added to cell cultures 80h after seeding. The drug exposure of cell cultures was 16h, with 0.1 μ M as final concentration.

Liquid N2 quenching

The falcon tube containing scraped cells (see above) was immersed into liquid nitrogen upon complete freezing of the sample and then slowly thawed in an ice bath. Afterwards the cells were washed twice in 10 ml of phosphate-buffered saline (PBS, pH 7.4), counted and the final volume was adjusted to obtain 30×10^6 cells into 5.4 ml PBS. Finally, the quenched cells were lysed by sonication to destroy the cell membrane and release intracellular metabolites.

Metabolites extraction for NMR analysis

Intracellular metabolites were extracted using a dual phase extraction procedure introduced by Bligh and Dyer in 1959 [36] with slight modifications. Adding 6 ml of cold methanol (-20°C) and 6 ml of chloroform to the original solution (5.4 ml) containing quenched cells, briefly a mixture of water, methanol and chloroform in the volume ratio of 0.9:1:1 was obtained, corresponding to a total volume of 17.4 ml. Afterwards, this mixture containing quenched cells was incubated for 20 min on ice and vortexed

frequently to facilitate the extraction. After the metabolites extraction step, cell extracts were centrifuged at 14000 rpm at 4°C for 20 min. This extraction procedure generated a two phase extract that can be described as follow: the aqueous upper phase contains water soluble intracellular metabolites, while apolar metabolites as lipid molecules are in the organic lower phase. Proteins and macromolecules are trapped in the layer between the two phases. The upper and lower phase were separated and carefully transferred into different falcon tubes. Eventually, solvents were completely removed from both fraction using a vacuum concentrator (aqueous phase) and under a gentle flow of N₂ gas (organic phase).

Successively, in this study hydrophilic (methanol extracts) and lipophilic (chloroform extracts) intracellular metabolites were analyzed by NMR spectroscopy.

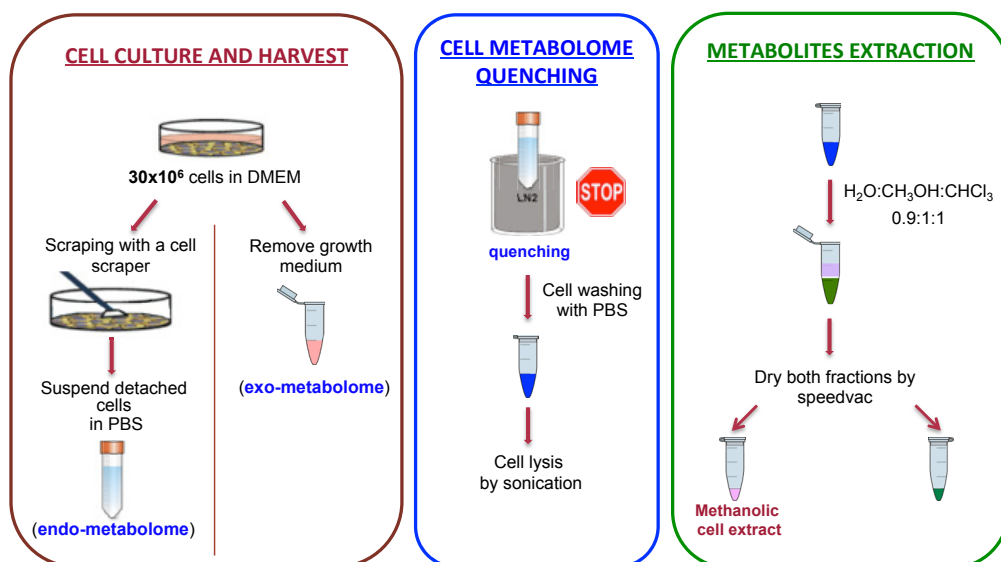


Fig.9 Illustration of the experimental protocol tested for NMR metabolomic analysis of
HTC-116 cancer cell lines

4.6 Experimental protocol optimization for NMR metabolomic analysis of HTC-116 cell lines

HTC-116 cells were grown in the absence or in the presence of treatments following the previously described culture conditions (see above) and testing a different amount of cells (15×10^6).

In order to improve the yield of the metabolites extraction a fundamental variation was introduced on the previous experimental protocol: after the culture medium removal, cells were washed with 15 ml PBS before cell scraping and quenching steps in order to avoid the loss of metabolites for cell leakage (Fig.10).

In addition, since this kind of cells grows as monolayer culture, four cell washing steps with PBS were needed to ensure the complete removal of residual culture medium.

Then, the last cell washing PBS was removed by aspiration, 5.4 ml of clean PBS were added to the culture dish and cells were mechanically scraped, finally ready for harvesting and quenching steps.

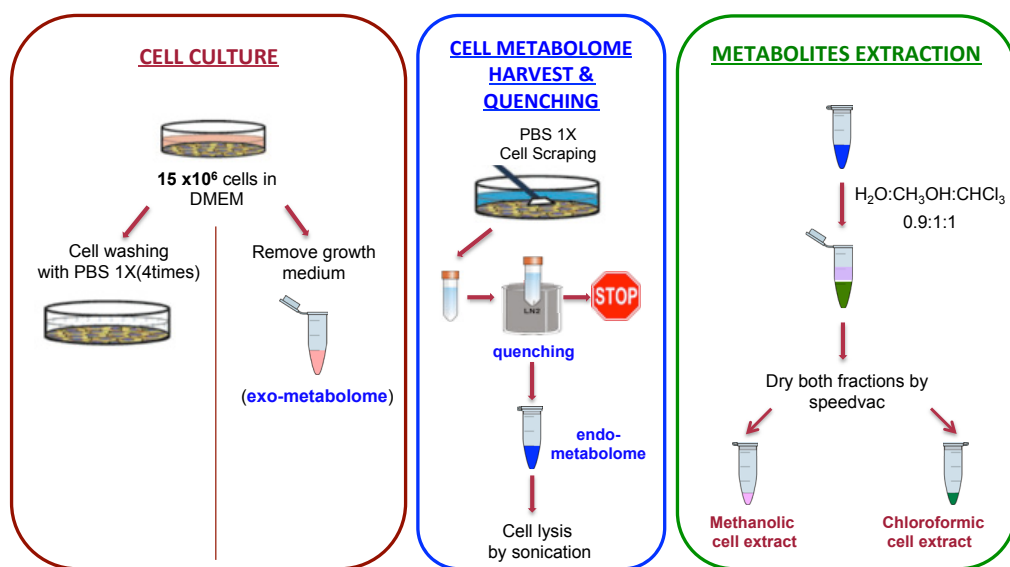


Fig.10 Illustration of the optimized experimental protocol tested for NMR metabolomic analysis of HTC-116 cancer cell lines.

Cell metabolome quenching

Two different quenching procedures were tested on the scraped cells: cold aqueous methanol ($-20^{\circ}C$) and liquid N_2 .

- ***Cold methanol quenching***

Cells were quenched using 6 ml of cold methanol ($-20^{\circ}C$) and 5.4 ml of ice-cold PBS. Next, cells were gently detached by scraping lightly the bottom of the culture dish using a cell lifter.

- ***Liquid N_2 quenching***

The falcon tube containing scraped cells was immersed into liquid nitrogen and then slowly thawing the sample in an ice bath. The detached cells were suspended in 10 ml of PBS in a falcon tube and then counted. Finally, the quenched cells were lysed by sonication to destroy the cell membrane and

release intracellular metabolites.

Metabolites extraction for NMR analysis

The solutions of quenched cells obtained from the two quenching procedures were separately suspended in 6 ml of chloroform (for methanol quenched cells) or 6 ml of methanol + 6 ml of chloroform (for liquid N₂ quenched cells). For both solutions, the metabolites extraction was performed using the same protocol as previously described. Samples were stored at -80°C until the NMR analysis.

4.7 NMR sample preparation

Prior to *endo*-metabolome NMR analysis, each aqueous cell extract was thawed and then reconstituted in 540 μl of D_2O together with 60 μl of a D_2O solution containing the sodium salt of (trimethylsilyl) propanoic-2,2,3,3- d_4 acid (TSP) , as an internal chemical shift reference (δ ^1H 0.00), to give a final concentration of 0.6 mM in the NMR tube. Samples were vortexed briefly and transferred into a 5-mm NMR tube.

The organic extracts were re-suspended in 700 μl of deuterated CDCl_3 containing 0.03% v/v tetramethylsilane (TMS) as chemical shift reference for organic solvent, then vortexed and transferred into a 5-mm NMR tube.

Before the *exo*-metabolome NMR analysis, 630 μl of the cell medium removed from the culture dish was added 70 μl of a D_2O solution containing TSP and transferred into a 5-mm NMR tube. Culture media samples were only briefly vortexed in order to minimize the bubbles formation attributable to high-abundance of serum proteins.

4.8 ^1H NMR Data acquisition

All NMR spectra were acquired at 300K on a Bruker Avance III 600 MHz ultrashielded spectrometer (Bruker Biospin GmbH, Rheinstetten, Germany) operating at 600,13 MHz for protons (14.09 Tesla) equipped with a double tuned cryo-probe (TCI) set for 5 mm sample tubes.

^1H NMR spectra of hydrophilic and lipophilic cell extracts were acquired using a 1D NOESY-presat pulse sequence (RD-90°-t-90°- t_m -90°-ACQ). A pre-saturation of the water resonance during the recycle delay is followed by a 90 degree pulse with an acquisition time of 2.726 s, a relaxation delay of 4 s, mixing time of 10 ms, receiver gain of 181, 128 scans, 128 K data points and a spectral width of 18029 Hz (30.041 ppm)

^1H NMR spectra of cell culture media were acquired using a 1D Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence (RD-90°- $\{\tau$ -180°- $\tau\}_n$ -ACQ) with presaturation for suppression of high molecular weight molecules signals on the basis of T_2 editing (300 μs delay and repeated 14 times). 128 free induction decays (FIDs) were collected into 128 K data points, using a spectral width of 12019 Hz (20.028 ppm), acquisition time of 3.067 s, with a 4 s relaxation delay between pulses, mixing time of 10 ms and receiver gain of 40.3. All samples were individually tuned, matched and shimmed. Prior to Fourier transformation, each free induction decay (FID) was zero-filled to 128 K points and multiplied by an exponential function equivalent to a 1.0 Hz line-broadening. The resulting spectra were phase- and baseline-corrected automatically using TOPSPINTM (Bruker Biospin) and the ppm scale was referenced towards the TSP peak at 0.00 ppm for aqueous solvents, the TMS peak at 0.00 ppm for organic solvent.

4.9 NMR Data Analysis and selection of experimental protocol

Despite mammalian cell metabolomics has being an increasing research field, the number of studies concerning quenching and extraction methods for adherent mammalian cells reported in the literature is still low. Different experimental protocols described in other papers were evaluated and adapted for HTC-116 cell lines NMR metabolomics study.

Initially we tested the metabolite extraction efficacy starting from 30×10^6 cells, washed with PBS after the scraping and quenching steps (as previously described in paragraph 3.5). Inspecting the 1D-NOESY spectrum (Fig.11) of the hydrophilic cell extract, it showed a quite low number of metabolite signals and the extraction yield resulted not sufficiently high. This NMR spectrum was acquired with 1024 scans number and took about 1 hour acquisition time.

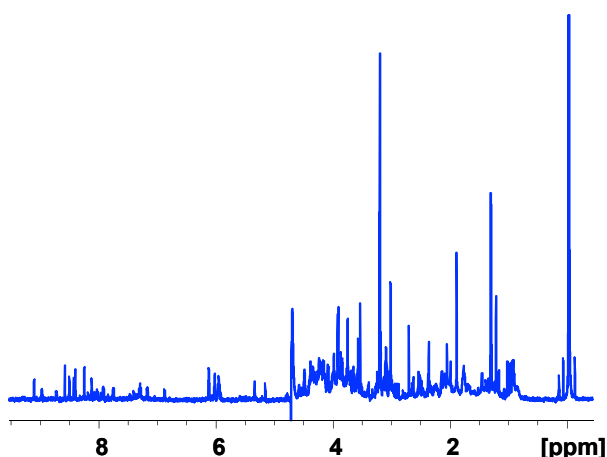


Fig.11 1D-NOESY spectrum of hydrophilic cell extract (control sample 30×10^6 cells, ns=1024, aq. t= 1h)

In order to improve the yield of metabolite extraction, we slightly changed the experimental protocol anticipating PBS cell washing before scraping and

quenching, as described in section 3.6. Starting from a lower number of cells (15×10^6), washed with PBS before the cell scraping and quenching steps and acquiring the 1D-NOESY spectrum (Fig.12) with only 128 number of scans, we noticed a considerable increase of the extraction yield. This was the proof that the PBS cell washing after the quenching step was the main responsible for the metabolites loss.

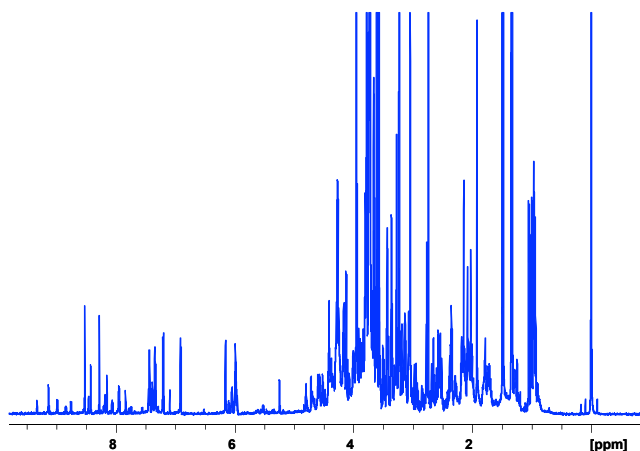


Fig.12 1D-NOESY spectrum of hydrophilic cell extract (control sample, 15×10^6 cells, ns=128, aq. t=15 min)

Moreover, using the same experimental conditions, we also tested the cold-methanol quenching method and compared it with the liquid nitrogen one. Accordingly to the literature describing experimental protocols in the case of adherent cells, the liquid nitrogen freezing represented the optimal way to stop HTC-116 cells enzymatic activity with respect to using methanol that instead induced a strong leakage of intracellular metabolites. As a result, we noticed that some leakage was anyhow unavoidable and that the selected experimental protocol involving the liquid N_2 quenching represented a good tradeoff among metabolome alteration, metabolites leakage and all analytical measurement difficulties in our study.

4.10 Pre-Processing of the NMR spectra and Multivariate Data Analysis

Before conducting multivariate data analysis, the NMR spectra had to be cleaned and corrected from problems that usually affect this type of data. The NMR regions between 12.72 and 9.43 ppm, between 4.75 and 4.615 ppm and from 0.8 to -10.32 ppm were removed because they only contain noise or the residual deuterated water signal as well as the reference signal at 0.00.

A preliminary explorative (unsupervised) data analysis was performed on the full features hydrophilic NMR data set, using only the samples from the same batch. The result is shown in Fig.13, where a PCA scores plot of batch B is illustrated.

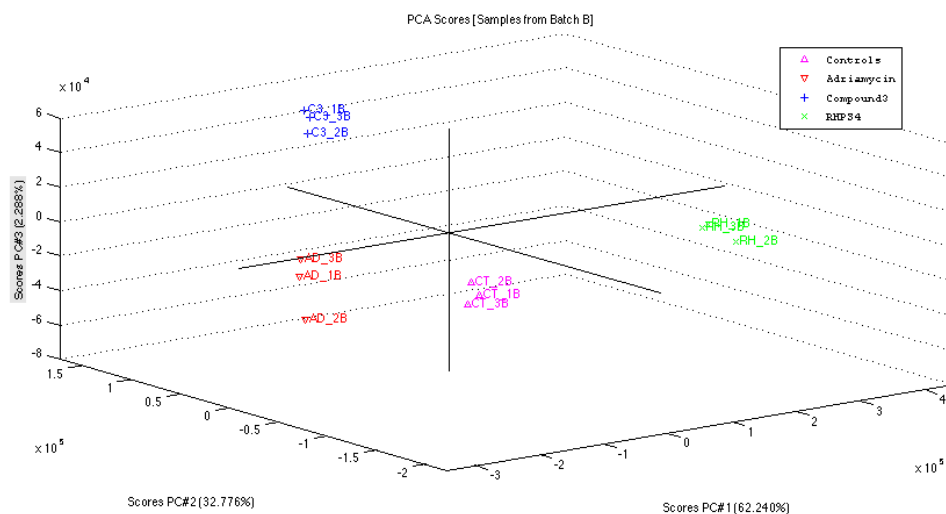


Fig.13 PCA model performed on the full features hydrophilic data. Samples of batches B are plotted.

Latentix 2.12 (www.latentix.com, Copenhagen, Denmark) was used to calculate the Principal Component Analysis (PCA) model.

In Figure 13 a clear separation among the control and the treatments groups is showed. Thereby, each treatment affects the HTC-116 cancer cells metabolism differently.

Moreover, the PCA model highlighted a considerable intra-batch experimental reproducibility, allowing us to make comparisons between samples groups and to understand the peculiarities of the treatments with respect to control samples.

In order to help pointing out how the different anticancer drugs act *in vitro* on the cellular metabolism, an overlay of the sample spectral average for each treatment from the batch B is shown in Figure 14 for a restricted region using different colors.

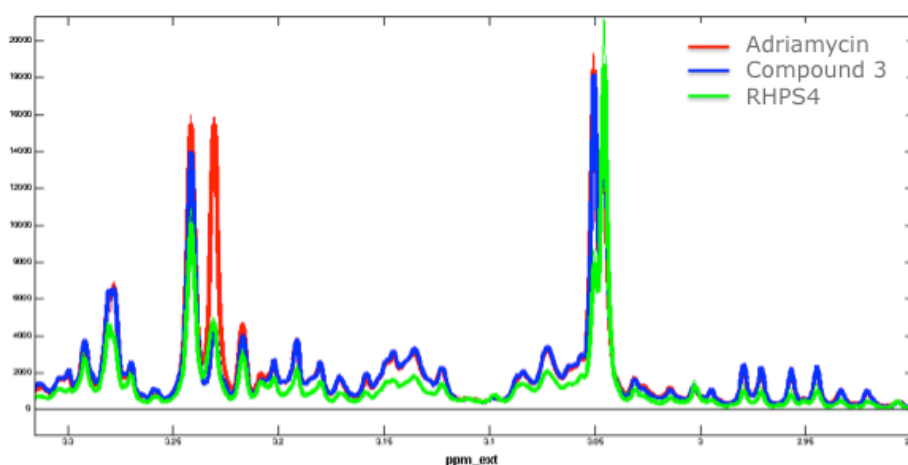


Fig.14 NMR spectra overlay of one sample for each drug treatment from the batch B. Zoom in the spectral region from 2.9 to 3.3 ppm.

Looking at the NMR spectrum of the sample treated with RHPS4, it is evident that globally, all spectral regions are interested with a substantial overall decrease in signal intensities. This presumably means that RHPS4 is

responsible of a strong cellular metabolism inhibition with respect to the other two drugs. Indeed, in the PCA model (Fig.13) RHPS4 treated group was separated from all the other groups by means of PC1, as proof of the RHPS4 diversity. Differently from Adriamycin and *Compound 3*, and in contrast with the overall tendency of lower metabolite production, a considerable increase in lactate formation was also showed for RHPS4 treatment.

Moreover, NMR signals such as those at 2.385, 3.04 and 8.35 ppm – unfortunately not yet assigned – are characteristic only for RHPS4, allowing us to understand that the treatment with RHPS4 determines the production of peculiar metabolites.

In order to check the inter-batch reproducibility, a Principal Component Analysis (PCA) was performed on the full features NMR data of the samples from the first (only control samples) and second batch (respectively A and B batch) (Fig.15).

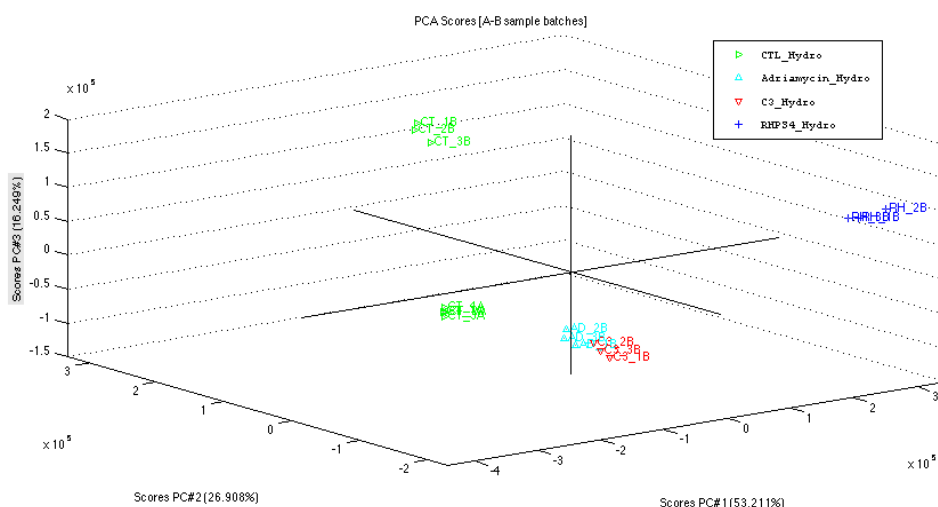


Fig.15 PCA model performed on the full features hydrophilic data. Samples of batches A and B are plotted.

The PCS scores plot of the resulting model showed an important separation between intra-batches control samples. This undesirable result might be due to different cellular generations of the same cell type which are used for preparing the samples of the two batches. Even though the cells used for the second batch are propagated from those used for the first one, their genotype and therefore their metabolome profile may be affected by generational changes.

From a data mining point of view, in order to overcome this problem and to minimize the difference between the different batches, a solution could be to use variable selection or data orthogonalization with respect to batch variation, but these approaches are tricky and may result in loss of information with respect to the original data.

Data normalization methods were also tested in this respect but they were not able to reduce the intra-batch variability.

Furthermore, since NMR spectra showed misalignments in chemical shift due to pH-sensitive peaks, the spectra were initially divided in equally sized (0.02 ppm) bins trying to minimize the inter-individual differences in peak positions. Principal Component Analysis (PCA) was performed on the binned data (Fig.16) for the hydrophilic extract samples again belonging to the first (A) and second batch (B).

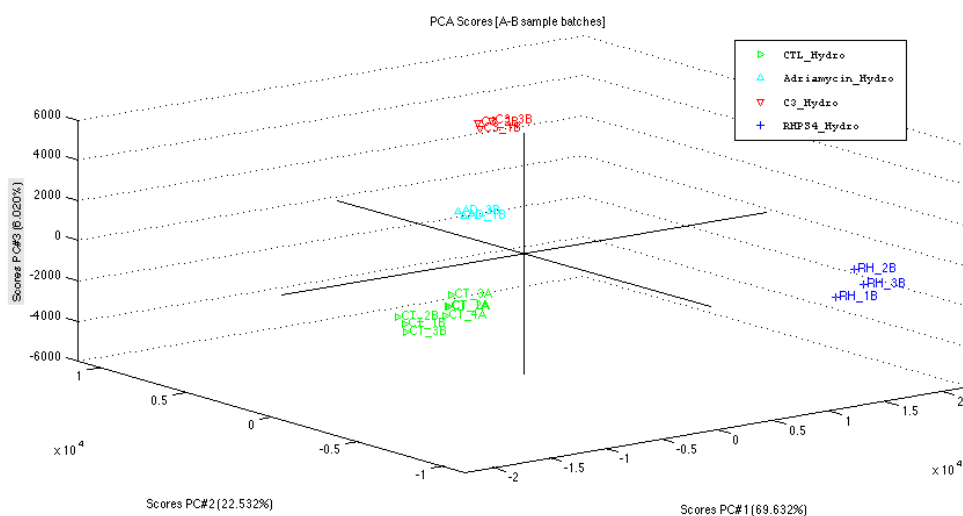


Fig.16 PCA model performed on the hydrophilic bin data (0.02 ppm).
Samples of batches A and B are plotted.

Comparing the PCA scores plot of this model with that of the previous one in Fig.15, the NMR spectra binning made the two control groups very comparable.

Thus, the spectral binning improved the grouping of the control samples of the two batches, indicating that a signal alignment is beneficial for improving the significance of the data. It must be highlighted that the binning method, although still widely used, always results in the loss of a considerable amount of information originally contained in the original spectra and should be avoided when other solutions that do not require drastic data reduction are available.

Therefore, in order to overcome the peak shift problem, the full resolution data matrix was corrected for errors in chemical shift misalignments concerning pH-dependent signals using the interval correlation optimized shifting algorithm (icoshift).

In the Fig. 17 an example of icoshift algorithm application is showed, zooming in the lactate quartet spectral region. Instead of bucketing method, icoshift does not involve a data reduction and the spectral resolution is maintained [37].

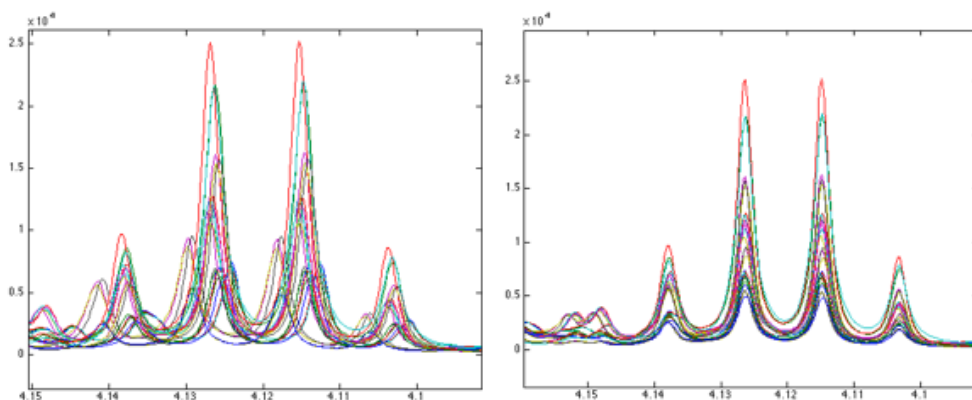


Fig.17 A comparison between raw (first plot) and icoshift aligned Lactate quartet (second plot)

Hence, all PCA models were performed on the aligned pareto-scaled data matrix. The pareto-scaling technique reduces the relative importance of large values (high intensities as for lactate) but keeps the data structure partially intact, assigning importance to the less intense signals, as for example those of the spectral aromatic region [38].

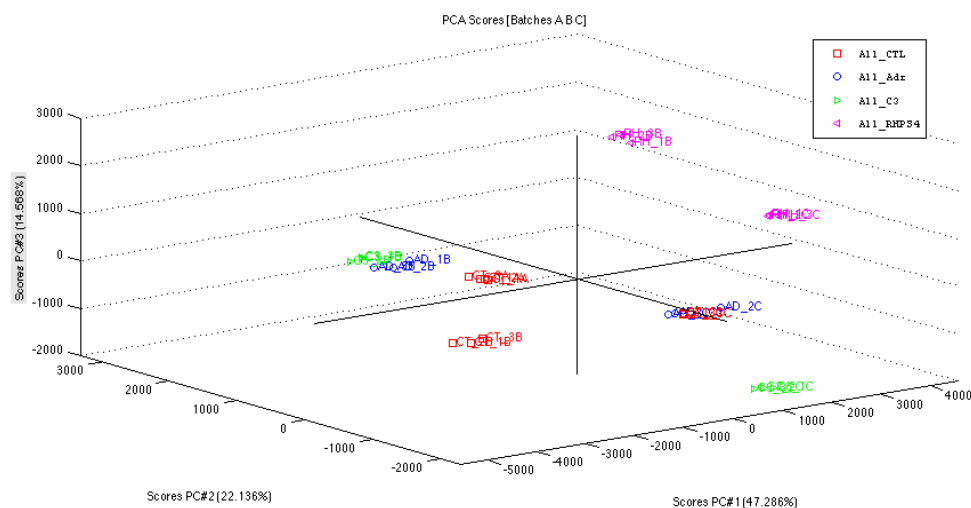


Fig.18 PCA model performed on the hydrophilic aligned data matrix. Samples of batches A - B and C are plotted.

Fig. 18 shows the scores plot of the PCA model performed on the aligned data matrix including samples from all batches; therefore also batch c that has both control samples and treated ones. It results clear that, even with a higher number of samples treated the same way, PC1 mainly describes the batch differences and the treated samples are well grouped only batch wise. Thus, the spectral data alignment was not sufficient to minimize the intra-batch differences, but it reduced the shift effect that was disturbing sample grouping. Once again the effect of RHP34 on the metabolome, described by PC1, is dominant and it even overcomes the batch effect.

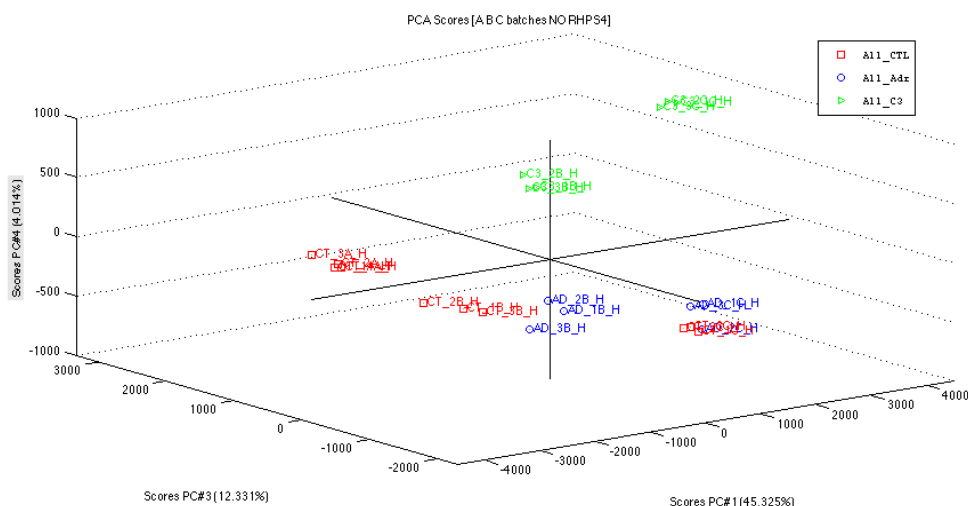


Fig.19 PCA model performed on the hydrophilic aligned data matrix. Samples of batches A - B and C without RHPS4 are plotted.

Recalculating the model without RHPS4 treated samples and plotting the PC score1 against PC score3 and 4, the samples resulted separated to a good extent on the basis of the anticancer drug treatment and not anymore of the different batches (Fig. 19).

It must be pointed out that PCA analysis has a solely explorative purpose and it is not supervised as for classification tools like PLS-DA or ECVA. Therefore, the non-perfect clusterizations obtained are anyhow indicative of a statistically robust difference among the different drug treatments. Hence, it can be stated that cell NMR metabolomics can be reliably used for studying the way the different anticancer drugs affect the cell metabolism, directly fulfilling the main hypothesis of the investigation. However, still a lot of work is required for interpreting the obtained results from a biological point of view and part of it has been initiated and described in the following chapter 4.11.

A preliminary explorative analysis approach was also carried out on the lipophilic extract and growth media samples data matrices. Both PCA models performed on the samples belonging to the same batches showed a significant difference among drug treatments (Fig. 20-21), but further investigations will aim to assess also the intra-batch variability for the lipophilic metabolome and the exo-metabolome.

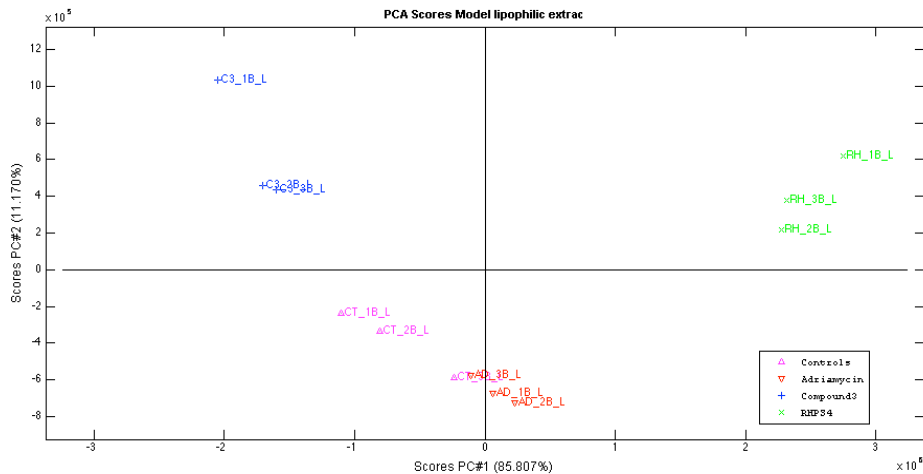


Fig.20 PCA model performed on the lipophilic non-aligned data matrix.
Samples of batches B are plotted.

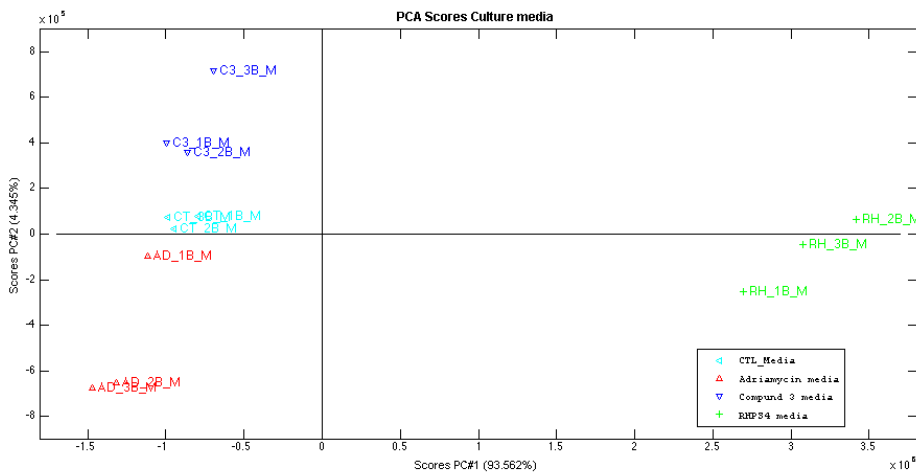


Fig.21 PCA model performed on the media non-aligned data matrix.
Samples of batches B are plotted.

It is interesting to note that the effect of RHPS4, described by PC1, is also dominant on the lipophilic metabolome and on the exo-metabolome. In particular, in order to understand which metabolites are interested by anticancer drug actions, we had a look at the NMR spectra overlay of lipophilic cellular extract (control and treated group) (Fig.22). It is clear that only the spectral region comprised from 1.3 ppm and 1.8 ppm (where lipoproteins use to resonate) is reasonably affected by drug treatment. This observation will be used to better describe the effect of each drug from a biological point of view. However, the exo-metabolome and the lipophilic cell extract profile still need to be further investigated.

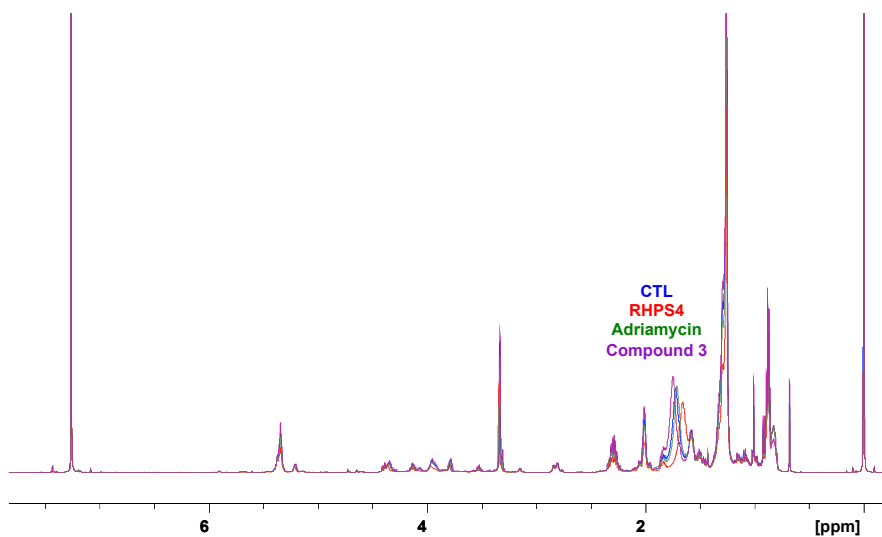


Fig.22 NMR spectra overlay of lipophilic cellular extract profile.

4.11 Metabolites identification

NMR hydrophilic metabolites identification in this project was achieved by comparison with the chemical shifts of the metabolites in the Human Metabolome Database (HMDB) (<http://www.hmdb.ca>) [39] and in the literature data [40] [41] [42].

Moreover, Statistical Total Correlation Spectroscopy (STOCSY) analysis was performed on the NMR (1D- NOESY) binned (0.02 ppm) data set of all samples belonging to the hydrophilic extracts (Fig. 23), for obtaining the correlations among the metabolite signals, using a threshold value $r > 0.9$ for correlation coefficient.

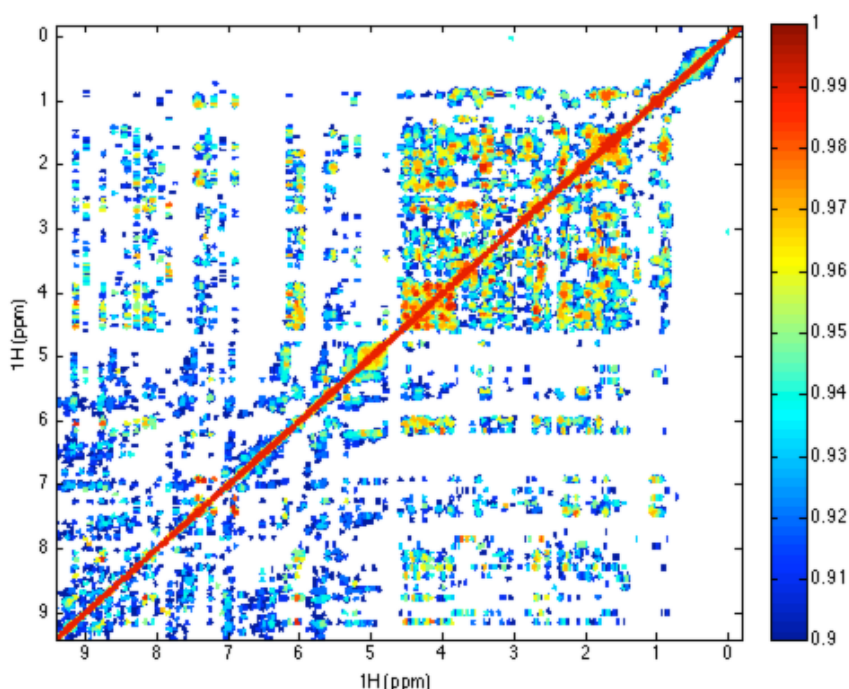


Fig.23 STOCSY correlation matrix plot. 1D-NOESY bin data (0.02ppm), hydrophilic extract samples were used to perform this analysis.

The resulting plot can be interpreted as a bidimensional NMR TOCSY spectrum where peaks on the diagonal are characterized by an autocorrelation value of 1. Highly correlated peak intensities represent protons belonging to the same molecule, whereas lower (or negative) correlations could lead to proton signals from different compounds but involved in the same metabolic pathway.

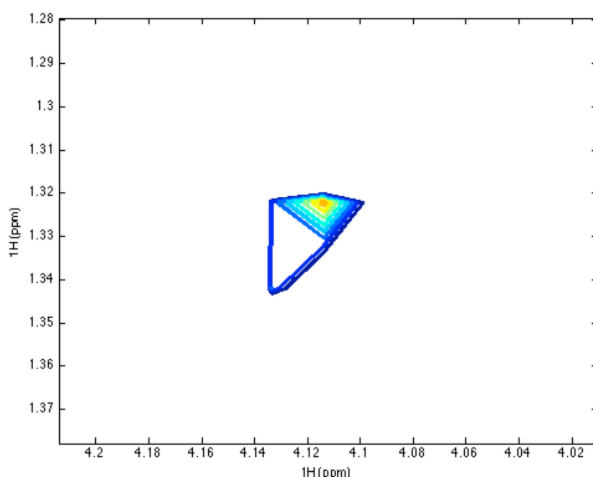


Fig.24 STOCSY zoom in the region showing the correlation crosspeak among lactate protons at δ_H 1.32 ppm and δ_H 4.115

For instance, a high correlation was observed among resonances δ_H 1.32 ppm and δ_H 4.115 (Fig.24), which corresponds to the lactate doublet and quartet respectively.

Hence, in Table 1 are presented the chemical shift and the signals multiplicity determined for each metabolite that have been unambiguously assigned.

Compound	Multiplicity	Shift (ppm) in D ₂ O
Acetate	s	1.92
Alanine	d	1.49
	q	3.80
D-Glucose	d	5.24 (α -anomer)
	dd	3.8-3.3 (overlap)
Glycine	s	3.575
Glutamate	m	2.03
	t	3.78
	m	2.35
	m	2.30
Glutamine	m	2.08
	t	3.8 (overlap)
	m	2.54
	m	2.57
Isoleucine	d	1.01
	t	0.96
Lactate	d	1.32
	q	4.115
Leucine	d	0.95
	d	0.93
Phenylalanine	d	7.34
	m	7.39
	m	7.44
	dd	3.95
	dd	3.048
	dd	3.235
Pyruvate	s	2.383
Threonine	d	1.33
	m	4.27
	d	3.62
Valine	d	1.05
	d	0.995

Tab.1 Chemical shift and signals multiplicity for HTC-116 cancer cell lines metabolites. Chemical shifts are reported with reference to TSP singlet resonance at 0.000 ppm, and multiplicity definitions are: s, singlet; d, doublet; dd, doublet of doublets; t, triplet; q, quartet; m, other multiplet.

4.12 Conclusions and future perspectives

Despite mammalian cell metabolomics is an increasing research field, the number of studies reported in the literature for adherent mammalian cells is still quite low. For a holistic understanding of cancer cells metabolome, it is crucial the choice of the right protocol for efficient harvesting, quenching and extraction of cellular metabolites. In this study we identified and optimized the experimental protocol for HTC- 116 human colon cancer cell lines for NMR metabolomics study.

From a general point of view, this project has relevant importance since colon carcinoma is the third most commonly diagnosed cancer in the world and the second most common cause of death from cancer.

The NMR analysis of metabolic profiles of human colon cancer cell lines aimed to provide important information on *in vitro* actions of drugs, pointing at their rapid incorporation into novel therapeutic settings. In particular, the PCA models calculated on the samples belonging to the same batch showed excellent results about the differentiation among the drug treatments and the intra-batch reproducibility.

Considering the exploratory characteristic of PCA analysis, which represents a non supervised approach to sample clustering, even the non-perfect groupings obtained are undoubtedly indicative of a statistically robust difference among the different drug treatments. Thus, it can be safely stated that cell NMR metabolomics represents a powerful tool for investigating how the different anticancer drugs affect the cell metabolism. This result fulfilled the main hypothesis of this investigation, but it only represents the starting point towards the understanding of cancer cell metabolism when anticancer drugs

are utilized.

In order to increase the statistical power of the study, we had to increase the number of investigated samples and, in order to do so, we collected samples from three cell production/treatment batches. The PCA models performed on the resulting data highlighted a substantial inter-batch variability. We tried to minimize this variation applying pre-processing methods, unfortunately without noticeable results.

Chemometric methods, such as any kind of variable selection tools and data orthogonalization could be used to minimize the batch effect, but a further study should be aimed at understanding the variability source and correct for it in order to guarantee that the selected protocol provides comparable results, avoiding the alteration of the original data.

In the next future we will try to go through the actions of the different drugs on the cellular metabolome (endo- and exo-metabolome), unequivocally assigning the metabolite NMR signals responsible for the differentiation among the treatments through advanced multivariate data analysis for biomarker profiling.

Eventually, we will map the profile of the identified metabolites into metabolic pathways to gain a biological overview of the mechanisms regulating drug-driven cell apoptosis and the selectivity with respect to cancer cells.

The interesting and exciting stimuli that I received during my PhD visiting period at the University of Copenhagen, animate my motivations for my future research efforts.

4.13 REFERENCES

- [1] Griffin J.L. Metabonomics: NMR spectroscopy and pattern recognition analysis of body fluids and tissues for characterisation of xenobiotic toxicity and disease diagnosis. *Curr Opin Chem Biol* **2003**, 7, 648–654.
- [2] Clayton T.A., Lindon J.C., Cloarec O., Antti H., Charuel C., Hanton G., Provost J.P., Le Net J.L., Baker D., Walley R.J., Everett J.R., Nicholson J.K. Pharmacometabonomic phenotyping and personalized drug treatment. *Nature* **2006**, 440, 1073–1077.
- [3] Robertson D.G. Metabonomics in toxicology: a review. *Toxicol Sci* **2005**, 85, 809–822.
- [4] Tang J. Microbial metabolomics. *Current Genomics* **2011**, 12, 391–403.
- [5] Cuperlovic-Culf M., Barnett D.A., Culf A.S., Chute I. Cell culture metabolomics: Applications and future directions. *Drug Discovery Today* **2010**, 15, 610–621.
- [6] Zhang A., Sun H., Xu H., Qiu S., Wang X. Cell Metabolomics. *OMICS A Journal of Integrative Biology* **2013**, 17, 10, 495–501.
- [7] Bai J., Wang M.X., Chowbay B., Ching C.B., Chen W.N. Metabolic profiling of HepG2 cells incubated with S(–) and R(+) enantiomers of anti-coagulating drug warfarin. *Metabolomics* **2011**, 7, 353–362.
- [8] Panopoulos A.D., Yanes O., Ruiz S., Kida Y.S., Diep D., Tautenhahn, R. Herrerias A., Batchelder E.M., Plongthongkum N., Lutz M., Berggren W.T., Zhang K., Evans R.M., Siuzdak G., Izipisua Belmonte J.C., *Cell Res.* **2012**, 22, 168–177.
- [9] Serkova N.J., Glunde K. Metabolomics of cancer. *Methods Mol Biol* **2009** 520, 273–295.
- [10] Jemal A., Bray F., Center M.M., Ferlay J., Ward E., Forman D. Global Cancer Statistics. *CA Cancer J Clin* **2011**, 61, 2, 69–90.
- [11] Teng Q, Huang W.L, Collette T.W, Ekman D.R, Tan C. A direct cell quenching method for cell-culture based metabolomics. *Metabolomics* **2009**, 5, 2, 199–208.

- [12] Sellick C.A., Hansen R., Maqsood A.R., Dunn W.B., Stephens G.M., Goodacre R., Dickson A. J. Effective Quenching Processes for Physiologically Valid Metabolite Profiling of Suspension Cultured Mammalian Cells. *Anal. Chem.* **2009**, 81, 174–183.
- [13] Wittmann C., Kromer J.O., Kiefer P., Binz T., Heinze E. Impact of the cold shock phenomenon on quantification of intracellular metabolites in bacteria. *Anal. Biochem.* **2004**, 327, 135–139.
- [14] Villas-Boas S.G., Moxley J.F., Akesson M., Stephanopoulos G., Nielsen J., High-throughput metabolic state analysis: the missing link in integrated functional genomics of yeasts. *Biochem. J.* **2005**, 388, 669–677.
- [15] Dietmair S., Timmins N.E., Gray P.P., Nielsen L.K., Krömer J.O. Towards quantitative metabolomics of mammalian cells: Development of a metabolite extraction protocol. *Anal. Biochem.* **2010**, 404, 155–164.
- [16] Lorenz M.A., Burant C.F., Kennedy R.T. Reducing Time and Increasing Sensitivity in Sample Preparation for Adherent Mammalian Cell Metabolomics. *Anal. Chem.* **2011**, 83, 3406–3414.
- [17] Duarte I.F., Marques J., Ladeirainha A.F., Rocha C., Lamego I., Calheiros R. Analytical approaches toward successful human cell metabolome studies by NMR spectroscopy. *Anal. Chem.*, **2009**, 81, 5023–5032.
- [18] Dettmer K., Nürnberger N., Kaspar H., Gruber M.A., Almstetter M.F., Oefner P.J. Metabolite extraction from adherently growing mammalian cells for metabolomics studies: optimization of harvesting and extraction protocols. *Anal Bioanal Chem*, **2011**, 399, 1127–1139.
- [19] Sellick C., Knight D., Croxford A., Maqsood A., Stephens G., Goodacre R., Dickson A. Evaluation of extraction processes for intracellular metabolite profiling of mammalian cells: matching extraction approaches to cell type and metabolite targets. *Metabolomics*, **2010**, 6, 3, 427–438.
- [20] Arcamone F., Cassinelli G., Franceschi G. Structure and physicochemical properties of adriamycin (doxorubicin). **1972** In: Int. Syrup. Adriamycin, pp. 9-22, CARTER, S. K., DIMARCO, A. and GniONE, M. (eds) Springer, New York.

- [21] Young R.C., Ozols R.F., Myers C.E. The anthracycline antineoplastic drugs. *New Engl. J. Med.*, **1981**, 305, 139-153.
- [22] Fornari F.A., Randolph J.K., Yalowich J.C., Ritke M.K., Gewirtz D.A. Interference by doxorubicin with DNA unwinding in MCF-7 breast tumor cells. *Mol Pharmacol*, **1994**, 45, 4, 649–56.
- [23] Momparler R.L., Karon M., Siegel S.E., Avila F. Effect of adriamycin on DNA, RNA, and protein synthesis in cell-free systems and intact cells. *Cancer Res*, **1976**, 36, 8, 2891–5.
- [24] Pommier Y., Leo E., Zhang H., Marchand C. DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chem. Biol.*, **2012**, 17, 5, 421-433.
- [25] Cosconati S., Rizzo A., Trotta R., Pagano B., Iachettini S., De Tito S., Lauri I., Fotticchia I., Giustiniano M., Marinelli L., Giancola C., Novellino E., Biroccio A., Randazzo A. Shooting for selective druglike G-quadruplex binders: Evidence for telomeric DNA damage and tumor cell death. *J. Med. Chem.*, **2012**, 55, 22, 9785-9792.
- [26] Biffi G., Tannahill, D., McCafferty, J., Balasubramanian S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **2013**, 5, 182–186.
- [27] Johnson J.E., Smith J.S., Kozak M.L., Johnson F.B. In vivo veritas: using yeast to probe the biological functions of G-quadruplexes. *Biochimie*, **2008**, 90, 1250-1263.
- [28] Henderson E., Hardin C.C., Walk S.K., Tinoco I.Jr, Blackburn E.H. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell*, **1987**, 51, 6, 899-908.
- [29] Cosconati S., Marinelli L., Trotta R., Virno A., Mayol L., Novellino E., Olson A.J., Randazzo A. Tandem application of virtual screening and NMR experiments in the discovery of brand new DNA quadruplex groove binders. *J Am Chem Soc.*, **2009**, 131, 45, 16336-7.
- [30] Di Leva F.S., Zizza P., Cingolani C., D’Angelo C., Pagano B., Amato J., Salvati E., Sissi C., Pinato O., Marinelli L., Cavalli A., Cosconati S., Novellino E., Randazzo A., Biroccio A. Exploring the chemical space of G-quadruplex binders: discovery of a novel chemotype targeting the human telomeric sequence. *J Med Chem.*, **2013**, 56, 23, 9646-54.

- [31] Leonetti C. , Scarsella M. , Riggio G. , Rizzo A. , Salvati E. , D'Incalci M. , Staszewsky L. , Frapolli R., Stevens M.F. , Stoppacciaro A. , Mottolese M., Antoniani B., Gilson E., Zupi G., Biroccio A. G-quadruplex ligando RHPS4 potenzia l'attività antitumorale di campotecine in modelli preclinici di tumori solidi. *Clin Cancer Res.*, **2008**, 14, 22, 7284-91.
- [32] Leonetti C., Amodei S., D'Angelo C., Rizzo A., Benassi B., Antonelli A., Elli R., Stevens M.F.G., D'Incalci M., Zupi G., Biroccio A. Biological activity of the G-quadruplex ligand RHPS4 (3,11-difluoro-6,8,13-trimethyl-8H-quino[4,3,2-kl]acridinium methosulfate) is associated with telomere capping alteration. *Mol Pharmacol.*, 2004, 66, 5, 1138–1146.
- [33] Salvati E., Leonetti C., Rizzo A., Scarsella M., Mottolese M., Galati R., Sperduti I., Stevens M.F.G., D'Incalci M., Blasco M., Chiorino G., Bauwens S., Horard B., Gilson E., Stoppacciaro A., Zupi G., Biroccio A. Telomere damage induced by the G-quadruplex ligand RHPS4 has an antitumor effect. *J Clin Invest.*, **2007**, 117, 11, 3236–3247.
- [34] Gavathiotis E., Heald R.A., Stevens M.F.G., Searle M.S. Recognition and Stabilization of Quadruplex DNA by a Potent New Telomerase Inhibitor: NMR Studies of the 2:1 Complex of a Pentacyclic Methylacridinium Cation with d(TTAGGGT)₄ . *Angewandte Chemie*, **2001**, 113, 24, 4885–4887.
- [35] Heald R.A., Modi C., Cookson J.C., Hutchinson I., Laughton C.A., Gowan S.M., Kelland L.R., and Stevens M.F.G. Antitumor polycyclic acridines. 8. (1) Synthesis and telomerase-inhibitory activity of methylated pentacyclic acridinium salts. *J Med Chem*, **2002**, 45, 590-597.
- [36] Bligh E.G., Dyer W.J. A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology*, **1959**, 37(8), 911-917.
- [37] Savorani, F.; Tomasi, G.; Engelsen, S. B. Icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. *J. Magn. Reson.* 2010, 202, 190–202.

- [38] Engel J., Gerretzen J., Szymanska E., Jansen J.J., Downey G., Blanchet L., Buydens L.M.C. Breaking with trends in pre-processing? *Trends in Analytical chemistry*, **2013**, 50, 96–106.
- [39] Wishart D.S., Jewison T., Guo A.C., Wilson M., Knox C., et al. HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic Acids Res.*, **2013**, Jan 1;41(D1):D801-7.
- [40] Bertini I., Xiaoyu H., Luchinat C. Global metabolomics characterization of bacteria: pre-analytical treatments and profiling. *Metabolomics*, **2014**, 10, 2, 241-249.
- [41] Govindaraju V., Young K., Maudsley A. Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR in Biomedicine*, **2000**; 13, 3, 129-153.
- [42] Duarte I.F., Marques J., Ladeirinha A.F., Rocha C., Lamego I., Calheiros R., et al. Analytical approaches toward successful human cell metabolome studies by NMR spectroscopy. *Analytical Chemistry*, **2009**, 81, 5023–5032.

Paper I

Malmendal A., Amoresano C., Trotta R., **Lauri I.**, De Tito S.,
Novellino E. and Randazzo A.

NMR Spectrometers as “Magnetic Tongues”:
Prediction of Sensory Descriptors in
Canned Tomatoes

Journal of Agricultural and Food Chemistry. 2011, 59,
10831–10838

NMR Spectrometers as “Magnetic Tongues”: Prediction of Sensory Descriptors in Canned Tomatoes

Anders Malmendal,^{*,†,‡} Claudia Amoresano,[§] Roberta Trotta,^{||} Ilaria Lauri,[⊥] Stefano De Tito,[⊥] Ettore Novellino,[⊥] and Antonio Randazzo^{*,||}

[†]Center for Insoluble Protein Structures, Interdisciplinary Nanoscience Center (iNANO) and Department of Chemistry, University of Aarhus, Langelandsgade 140, DK-8000 Aarhus C, Denmark

[‡]Department of Biomedical Sciences, Faculty of Health Sciences, University of Copenhagen, DK-2200 Copenhagen N, Denmark

[§]Loghia Market Research, Piazza Immacolata 12, 80129, Napoli, Italy

^{||}Dip. Chimica delle Sostanze Naturali and [⊥]Dip. Chimica Farmaceutica e Tossicologica, Università degli Studi di Napoli “Federico II”, via D. Montesano 49, 80131 Napoli, Italy

 Supporting Information

ABSTRACT: The perception of odor and flavor of food is a complicated physiological and psychological process that cannot be explained by simple models. Quantitative descriptive analysis is a technique used to describe sensory features. Nevertheless, the availability of a number of instrumental techniques has opened up the possibility to calibrate the sensory perception. In this frame, we have tested the potentiality of nuclear magnetic resonance spectroscopy as a predictive tool to measure sensory descriptors. In particular, we have used an NMR metabolomic approach that allowed us to differentiate the analyzed samples based on their chemical composition. We were able to correlate the NMR metabolomic fingerprints recorded for canned tomato samples to the sensory descriptors bitterness, sweetness, sourness, saltiness, tomato and metal taste, redness, and density, suggesting that NMR might be a very useful tool for the characterization of sensory features of tomatoes.

KEYWORDS: NMR, multivariate analysis, sensory analysis, canned tomatoes

■ INTRODUCTION

It is crucial to know consumers' expectations, habits, and preferences to ensure product success on the market. Brand, label information (such as geographic origin, technology, etc.), price, packaging, factory image, product concept, and effective communication are all critical factors. However, when the consumer decides whether to buy the product again or not, success is tightly connected to the products' features.

It is therefore extremely important to understand how much consumers' preferences are driven by differences in sensory features between products. Traditional consumer research helps determine acceptable versus unacceptable. It is helpful when an overall, synthetic understanding of the product's acceptance is needed. However, it is not of any help when an explanation, in terms of sensory descriptors, is needed to provide R&D with technical information useful to enhance product features. Such information can only be provided through analytical products evaluation, of which consumers are not capable.

A detailed sensory description, in fact, requires the ability to decompose each sensory feature, requires selective attention, and thus requires people specifically trained to the application of sensory analysis (quantitative descriptive analysis, QDA).¹ Sensory analysis is a discipline through which the sensory analyst evokes, measures, analyzes, and interprets human responses to stimuli as perceived through the senses. Human sensory tests are regularly employed in the food and beverages industries, and they are sometimes integrated by a number of techniques, including

the electronic nose² and the electronic tongue.² The most common types of sensors used are based on electrochemical techniques, such as potentiometry and voltammetry.^{3–5} Other sensing methods include optical⁶ and acoustic techniques.⁷ Furthermore, techniques like mass spectrometry (MS)⁸ and gas chromatography (GC)⁹ have also been used. ¹H nuclear magnetic resonance (NMR) spectroscopy also has been used to investigate the taste of wine.¹⁰ Here, we investigate the utility of ¹H NMR as a tool to analyze the taste of canned tomato without any other chemical analysis.

■ MATERIALS AND METHODS

Materials. Eighteen canned tomato products of different brands were purchased in different markets in the city of Napoli (Italy) (Table S1 in the Supporting Information).

Sensory Assessment. A panel of trained 12 assessors (six females and six males) was selected based on the ability to recognize, describe, and quantify basic tastes, odors, and texture properties. The panel developed a specific profile protocol for QDA containing 14 descriptors: redness, synaeresis, dimension, residual peel, consistency, density, tomato flavor, saltiness, sourness, bitterness, sweetness, tomato taste, cooked taste, and metal taste. Descriptors were evaluated on a continuous, unlabeled, 0–10 intensity scale and then turned into numeric variables

Received: June 11, 2011

Accepted: September 23, 2011

Revised: September 22, 2011

Published: September 23, 2011

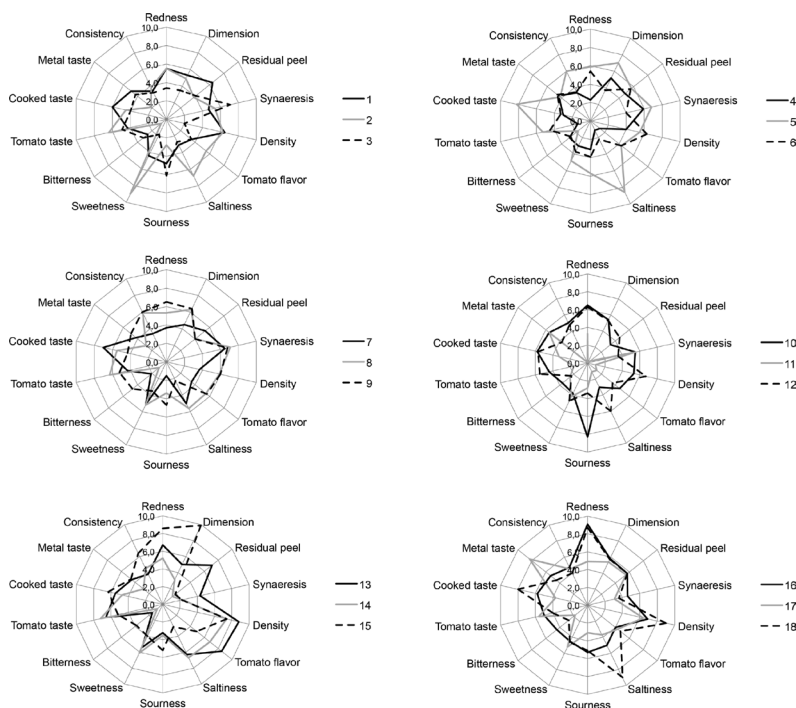


Figure 1. Spider web plot of the sensory descriptors for the 18 tested samples. The mean QDA parameters are listed in Table S2 in the Supporting Information.

(a number between 0 and 10). Three replicates per sample were performed, to minimize random errors (each subsequent replicate after 1 week from the previous one). The 18 samples were presented blinded in a flat plastic plate uncooked and at a controlled temperature (30 °C). A maximum of three samples were presented during each session according a balanced rotation plan.

Sample Preparation for ^1H NMR Analysis. Each sample was blended and centrifuged at 2200g for 30 min. Four aliquots (500 μL) of supernatant of each sample was diluted with 100 μL of D_2O and analyzed independently. No buffer was used.

Chemicals and Reagents. Deuterium oxide (D_2O , 99.9 atom %) was purchased from Cambridge Isotope Laboratories, Inc.

NMR Spectrometry. ^1H NMR spectra were acquired at 25 °C with a 700 MHz Varian Unity Inova spectrometer using a 5 mm $\{^1\text{H}/^{13}\text{C}/^{15}\text{N}\}$ triple resonance probe. The ^1H NMR measurements were carried out with 128 transients and 16K complex data point. The recycle time was set to 5 s, and a 45° pulse angle was used. The water signal was suppressed using presaturation.

NMR Data Reduction and Processing. The spectra were processed using iNMR (www.inmr.net). An exponential line-broadening of 0.5 Hz was applied to the free-induction decay prior to Fourier transformation. All spectra were referenced relative to external sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS), phased, and baseline corrected. Four aliquots of each product were studied by one-dimensional ^1H NMR. In total, 72 spectra were acquired. The spectra were aligned by correlation optimized warping¹¹ using $m_p = 50$ and $n_p = 2$. Data reduction was accomplished by dividing the spectrum into 0.01 ppm regions (bins) over which the signal was integrated to obtain the signal intensity. The region around the residual

water signal (5.0–4.7 ppm) was removed in order not to compromise the analysis. The high- and low-field ends of the spectrum, containing no signal, were also removed (i.e., leaving data between 9.5 and 0.5 ppm). At the end, a total of 870 variables were analyzed for each spectrum. The integrals were normalized to a total intensity to suppress trivial separation based on variations in the amount of sample.

The dendrograms describing the sensory analysis were based on unscaled sensory data. The NMR-based dendrograms were based on PLS-DA scores of VAST scaled¹² NMR data calculated using Simca-P 11.5 (Umetrics, Umeå, Sweden) as input. In VAST scaling,¹² each region/bin is divided by the average standard deviation of the integral of that region within each product. This scaling reduces the weight of random variations between “identical” samples, and the analysis is not biased toward compounds present at high concentrations. The number of axes for the PLS-DA model was determined by leave one out cross-validation, where all of the samples from each of the 18 products were left out for one product at a time to determine the quality of the model. The model used was estimated using all 18 products. Hierarchical cluster analysis (HCA) was then carried out using complete linkages in R (<http://www.r-project.org>) by using the Euclidean distance between the PLS scores for each product.

Principal component analysis (PCA) was carried out on unscaled sensory data. VAST-scaled¹² NMR data were used. The PCA was performed using Simca-P 11.5 (Umetrics). The number of principal components (PCs) was determined by leave one out cross-validation as described above. To test which PCs that varied significantly between products, the PC scores for the NMR data were subjected to one-way analysis of variance using sequential Bonferroni correction for multiple testing (significance level, 0.05). The fact that the variations between the samples from the

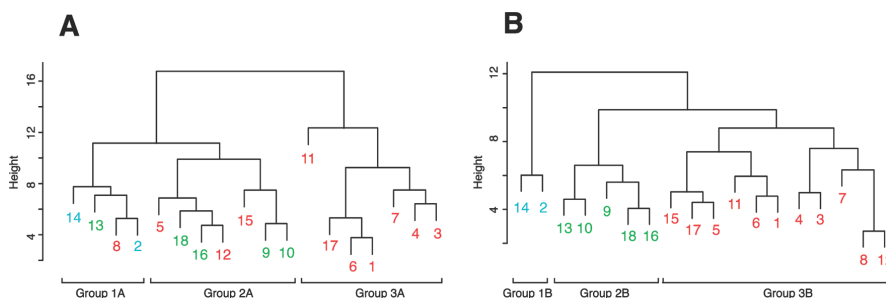


Figure 2. Dendrograms showing the similarities between products based on (A) QDA and (B) NMR. Products falling within the same group in the NMR classification are indicated with the same color.

same can were taken as the variation within the product might result in an overestimation of the significance. Standard errors (SEs) were calculated as $SE = SD/N^{1/2}$ where SD is the standard deviation and N is the number of samples from that product.

Orthogonal projection to latent structures, OPLS, separates the variance in x correlated with y (y -predictive) with the orthogonal (noncorrelated; y -orthogonal) variance.¹³ In contrast to regular PLS, a single y will result in only one predictive component. OPLS was carried out using each sensory descriptor as the y -variable. Data were scaled to obtain unit variance and then centered. OPLS was performed using Simca-P 12.0 (Umetrics). Cross-validation was obtained as described above. Markers for the sensory descriptors were identified from the NMR signals that showed a strong correlation ($R^2 > 0.5$) with the OPLS predictive scores for the sensory descriptors.

RESULTS AND DISCUSSION

Sensory Analysis. QDA mean results are reported in Figure 1. To group products sharing similar sensory features, HCA was performed on QDA means. The resulting dendrogram is shown in Figure 2A. Three main groups were identified, consisting of products 14, 13, 8, and 2 (group 1A); products 5, 18, 16, 12, 15, 9, and 10 (group 2A); and products 11, 17, 6, 1, 7, 4, and 3 (group 3A).

PCA was also performed on the same data set (Figure 3). Two PCs accounting for 60% of the variation were identified. A plot of their scores (Figure 3A) shows the positioning of the products according to their sensory attributes and allowed the identification of the most important sensory descriptors for products differentiation. This analysis indicates that the groups identified by the HCA share the same features and that there is no strong separation between the different groups identified. According to the loading plot (Figure 3B), the transition from the upper-left corner to the bottom-right corner of the map shows the simultaneous decrease of the bitterness and metal taste and increase of the sweetness and saltiness. Tomato flavor, saltiness, and tomato tastes are positioned on the bottom-right side of the map. Redness, consistency, dimension, density, residual peel, sourness, and cooked taste are positioned in the upper-right quadrant. In general, products belonging to group 1A are characterized by sweetness, by tomato taste and saltiness, and by tomato flavor. Group 2A is instead characterized by a more marked redness and sourness. On the other hand, group 3A is characterized by bitterness and metal taste, having a light

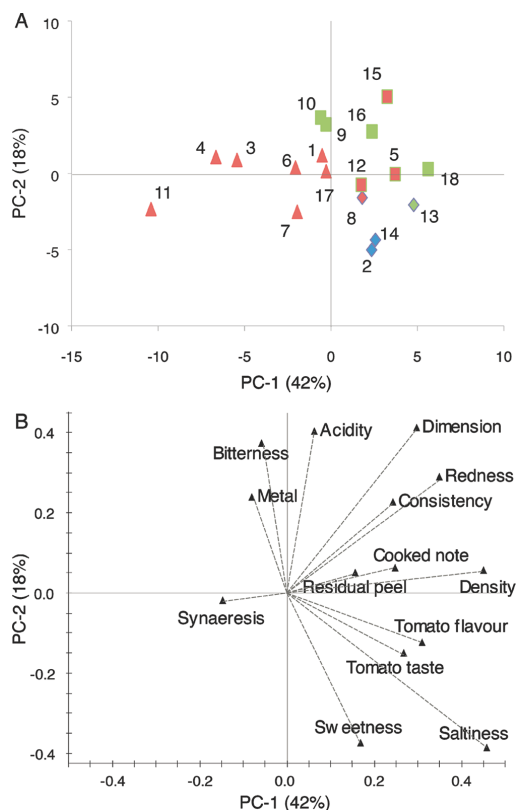


Figure 3. Score (A) and loading (B) plots of the PCA performed on sensory data. Products are colored according to NMR HCA analysis in Figure 2B. Note that none of the sensory descriptors are well described by this PCA model ($|R| > 0.5$ for all descriptors).

redness. However, none of the descriptors shows a high correlation ($|R| > 0.5$) with the model (Figure 3B).

To characterize the correlations between different sensory descriptors, the correlation coefficients were calculated (Table 1).

Table 1. Correlation Coefficients ($|R| > 0.5$) between Sensory Descriptors^a

	redness	dimension	synaeresis	density	tomato flavor	saltiness	sourness	sweetness	bitterness	tomato taste	cooked taste	consistency
redness		0.60		0.76							0.56	0.55
dimension	0.60			0.55							0.51	0.87
synaeresis				-0.60								
density	0.76	0.55	-0.60		0.71	0.57				0.59		0.56
tomato flavor				0.71		0.52				0.86		
saltiness				0.57	0.52			0.58	-0.52	0.57	0.71	
sourness									0.54			
sweetness						0.58			-0.69	0.52		
bitterness						-0.52	0.54	-0.69				
tomato taste				0.59	0.86	0.57		0.52				0.58
cooked taste	0.56	0.51				0.71						
consistency	0.55	0.87		0.56						0.58		

^a Sensory descriptors showing $|R| < 0.5$ to all other sensory descriptors are excluded.

We can see, for example, a strong negative correlation of sweetness with bitterness but not with metal taste or sourness, as suggested by the loadings plot (Figure 3B).

NMR Analysis. The same products tested in the QDA were analyzed by NMR. The superimposition of two representative ¹H NMR spectra is reported in Figure 4. It should be noticed that for each product, all of the NMR samples were taken from the same can. The data might thus underestimate the spread of the chemical properties within each product. Analogously to the HCA performed on sensory data, the HCA analysis performed on the NMR data revealed three main groups (Figure 2B): 14 and 2 (group 1B); 13, 10, 9, 18, and 16 (group 2B); and 15, 17, 5, 11, 6, 1, 4, 3, 7, 8, and 12 (group 3B). Despite the fact that the two HCAs refer to data collected by very different analytical techniques, it can be seen that there is a good global agreement between the different measurements: All products of group 1B (products 2 and 14) are also present in group 1A, all products except one in group 2B (products 9, 10, 16, and 18) are also present in group 2A, and all products in group 3A (products 1, 4, 6, 7, 11, and 17) are also present in group 3B.

PCA has also been performed on the NMR data set. Fourteen PCs were identified, of which PC1, PC2, and PC3 vary significantly between the different products. These three PCs account for 57% of the variation. The general distribution of the products in the score plots (Figure 5A,B) in a way recall the one observed in the sensory data set (Figure 3A). For example, considering the PC1–PC2 plot, products 3 and 4 are mapped close to each other and, at the same time, far away from the products 9, 10, 16, 15, and 18. Similarly, these latter samples are far way from products 2 and 14. Finally, the products 1, 6, 7, 8, 12, and 17, which were placed in the very center of the plot of the sensory data (Figure 3A), are placed in the center of NMR PC1–PC2 plot as well. As judged from the loading plots (Figure 5C–E), the first PC describes the distribution of the samples based on their sweetness. In fact, negative values can be observed for signals belonging to sugars like saccharose and α - and β -D-glucose. At the same time, positive correlations can be observed for signals belonging to bitter amino acids like tyrosine, phenylalanine, tryptophane, and isoleucine (see Chemical Signatures of Sensory Descriptors). The noisy look of the second PC describes the formation of sharper NMR signals due to a decrease in viscosity. The third PC seems instead related to an increase of saccharose, isoleucine, and acetate and a decrease of tyrosine, α -D-glucose, malate, and glutamate.

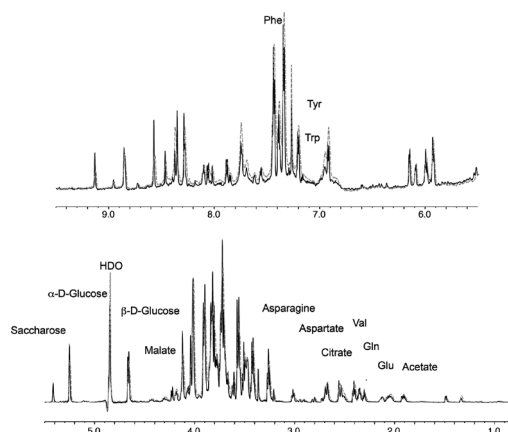


Figure 4. Annotated ¹H NMR spectra of two typical canned tomato samples. Product 2 (solid black line) is characterized by the presence of saccharose (see signals at 5.41 ppm) and a low viscosity, while product 15 (dashed gray line) is characterized by the absence of saccharose and a high viscosity. Note that the y-axis scale of the upper panel is increased 40× compared to the lower.

The loading plots also contain a number of signals that could not be assigned unambiguously.

Prediction of Sensory Descriptors. Although it is encouraging that there are similarities in the structures of the sensory and NMR data, the important question is how well the sensory descriptors can be predicted by NMR. To resolve that question, we made predictive models for the different sensory descriptors using orthogonal-projection to latent structures, OPLS.¹³ Using this protocol, we were able to get good predictions [$Q^2(\text{cum}) > 0.5$] for bitterness, redness, density, and metal and tomato taste (Table 2). After inspection of the remaining models and identification of outliers in those, we were able to get good models for all but two brands for saltiness, sweetness, and sourness (Table 2). In five cases out of six, the removed products showed extreme values for saltiness (2 of 2), sweetness (2 of 2), and sourness (1 of 2). It thus seems that these extra strong features depend on other factors than those under more normal conditions. The remaining

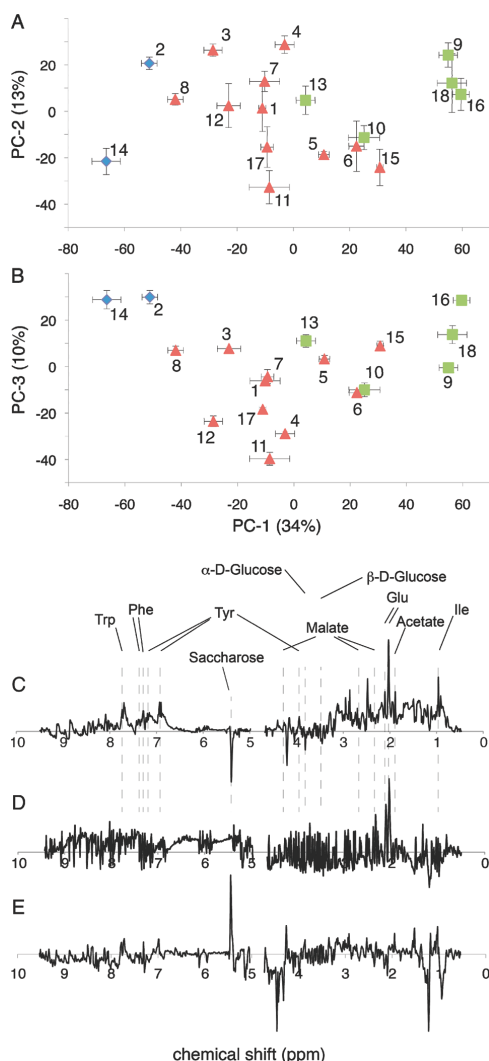


Figure 5. Score (A and B) and loading (C–E) plots of the PCA performed on NMR data. Panels A and B show the PC1–PC2 and PC1–PC3 score plots, and panels C–E show the PC1–PC3 loadings. Products are colored according to NMR HCA analysis in Figure 1B. Error bars correspond to one SE ($SE = SD/N^{1/2}$).

descriptors were related to the physical rather than chemical properties of the products.

Chemical Signatures of Sensory Descriptors. To determine the chemical components responsible for a given sensory descriptor, we have looked for all possible correlations between the NMR signals and the analyzed sensory descriptors using OPLS models. Specifically, the origin of signals displaying correlation above $R^2 > 0.5$ with the OPLS scores for the sensory descriptors with a $Q^2 > 0.5$ was identified. In this procedure, a multitude of

Table 2. Description and Statistical Summary of the OPLS Models Constructed Based on NMR Data

variable	A^a	N^b	$R^2X(\text{cum})^c$	$R^2Y(\text{cum})^c$	$Q^2(\text{cum})^d$
bitterness	7	66	0.70	0.99	0.87 ^e
redness	5	66	0.65	0.98	0.86 ^e
density	2	66	0.37	0.80	0.68 ^e
metal taste	1	66	0.31	0.85	0.67 ^e
tomato taste	2	66	0.42	0.87	0.58 ^e
saltiness	1	66	0.27	0.71	0.33
products 5 and 18 excluded	5	58	0.66	0.99	0.91 ^e
sweetness	2	66	0.40	0.84	0.30
products 2 and 3 excluded	7	58	0.72	0.99	0.78 ^e
tomato flavor	2	66	0.34	0.74	0.23
residual peel	1	66	0.30	0.56	0.14
consistency	2	66	0.43	0.73	0.07
	2	62	0.37	0.79	0.26
sourness	0	66	0.23	0.31	0.04
products 3 and 10 excluded	5	58	0.65	0.96	0.83 ^e
syneresis	1	66	0.29	0.59	0.02
dimension	0	66	0.21	0.36	0.02
cooked taste	0	66	0.21	0.32	−0.01

^a A number of orthogonal components. ^b Number of samples included in the model. ^c $R^2X(\text{cum})$ and $R^2Y(\text{cum})$ = the cumulated fraction of the variance in the parameter explained by the model. ^d $Q^2Y(\text{cum})$ = the cumulative predicted fraction of the variation of the parameter as determined by cross-validation. ^e $Q^2Y(\text{cum})$ values above 0.5 are considered as good predictors.

chemical components were identified for several of the sensory descriptors¹⁴ (see also Table 3).

It was possible to identify only very few compounds that have a relationship with sweetness. Particularly, sweet perception was positively correlated with saccharose (5.41 ppm) in spite of its low concentration, whereas it was negatively correlated with tyrosine (H- α 3.94 ppm), which is a known bitter amino acid. No correlation was found with citrate, while a negative correlation with the malate signal at 4.29 ppm was found. This is an interesting result since malate and citrate seem to have very similar sensory properties (see below).¹⁵ The characteristic sweet–sour taste of tomato and its overall flavor intensity are mainly due to reducing sugars, free acids, and free amino acids, minerals, and volatile substances. Overall, the character and intensity of taste are greatly affected by the salts present and by the buffer effect of the various cations and anions. About 50% of the dry matter in tomatoes is made of sugars, primarily glucose and fructose. There is frequently saccharose as well, but its quantity rarely exceeds 0.1% of the fresh mass.^{16,17} It is interesting to note that the sensation of sweetness cannot solely be explained by the sugar content. In fact, Jones and Scott did not find a close correlation between sugar content and sweetness.¹⁸ Similarly, Watada and Aulenbach did not find correlation between sweetness and dry matter content either.¹⁹ All of this means that other components affect the perceived sweetness. Interestingly, Stevens and co-workers found a relationship between the sensation of sweetness and the glucose/citric acid interaction.²⁰ Particularly, they have found that glucose affects sweetness more than fructose with high citric acid concentration. Furthermore, when the sugar concentration is low, citric acid reduces perceived sweetness, while with high sugar concentration, it increases sweet perception. It has been estimated that the relative composition in

Table 3. Correlation between Chemical Substances and OPLS Models for Sensory Descriptors^a

	redness	metal taste	saltiness	sourness	sweetness	bitterness	tomato taste
acetate (1.91)						+	
aspartate (2.68, 2.80)				+		+	
asparagine (2.87, 2.95)				+			
citrate (2.53, 2.66)				+			
glucose (3.49, 3.82)		—		—		—	
glutamate (2.05, 2.16, 2.32)				+		+	
glutamine (2.14, 2.45)				+			
isoleucine (0.93, 1.26, 1.46, 1.98)		+	—	+		+	—
malate (2.37, 2.66, 4.29)		+	—	+	—	+	—
phenylalanine (7.30, 7.40)		+		+		+	
saccharose (5.41)		—			+	—	
tryptophan (7.17, 7.29, 7.75)	+			+		+	
tyrosine (3.94, 6.90, 7.18)	+	+		+	—	+	
valine (2.53)				+		+	

^a + and — signs indicate positive and negative correlations, respectively. Chemical shift values (ppm) of the used signal are reported in brackets.

glucose, fructose, and citric acid can explain about 80% of the variation in sweetness.²⁰

In contrast to sweetness, bitter taste was negatively correlated with glucose and saccharose signals and positively correlated with a number of bitter amino acids²¹ like isoleucine (H- β , H- γ 1, H- γ 2, and Me- γ at 1.98, 1.46, 1.26, and 0.93 ppm, respectively), tryptophan (H4, H5, and H6 at 7.75, 7.17, and 7.29 ppm, respectively), tyrosine (H3/H5 and H2/H6 at 6.90 and 7.18 ppm, respectively), valine (H- β at 2.53 ppm), and phenylalanine (H3/H5 and H2/H6 at 7.40 and 7.30 ppm, respectively). The correlation with glucose suggests that even if glucose did not correlate with sweet taste, it has a strong masking effect on the bitter taste. Sweetness and bitterness show a relatively strong anticorrelation ($R = -0.69$; Table 2). Interestingly, bitter taste was also positively correlated with glutamate signals (H- β 1, H- β 2, and H- γ at 2.52, 1.62, and 32 ppm, respectively), acetate (1.91 ppm), and malate (2.37 and 2.66 ppm), all compounds that do not possess a bitter taste themselves. However, the taste-enhancing effect of the glutamic acid, one of the most abundant amino acid in tomato, was proven,^{22–24} and we cannot exclude a similar effect also for acetate and malate.

The sour taste of tomato can be ascribed mainly to the organic acids, rather than to the hydrogen ion concentration. Organic acids form more than 10% of the dry content of tomatoes.^{25,26} The two main acidic components are citric and malic acid, where malic acid is more sour than citric acid even if present in lower concentration. In our case, we found that sourness is positively correlated with both of these components. Moreover, it is known that sourness is also affected by the presence of free amino acids.²⁷ We found positive correlations with amino acids having taste-enhancing properties like glutamate, glutamine, aspartate, and asparagine and with amino acids having a bitter taste like tryptophan, tyrosine, phenylalanine, valine, and isoleucine. Interestingly, sour taste was negatively correlated to the presence of α - and β -D-glucose. All of these data strongly suggest that sour taste is closely correlated to bitter taste. As shown in Table 2, the correlation coefficient between the two was 0.54 in this study.

Furthermore, tomato taste and saltiness were all positively correlated (Table 2) and were negatively correlated to isoleucine (H- γ and Me- γ at 1.26 and 0.93 ppm, respectively) and malate (4.29 ppm). Metal taste, instead, had positive correlations with

bitter amino acids like isoleucine (H- γ and Me- γ at 1.26 and 0.93 ppm, respectively), tyrosine (H3/H5 and H2/H6 at 6.90 and 7.18 ppm), and phenylalanine (H3/H5 and H2/H6 at 7.40 and 7.30 ppm, respectively). On the other hand, a negative correlation was evident with the signal belonging to α - and β -D-glucose (3.82 and 3.49 ppm) and saccharose (5.41 ppm). Interestingly, metal taste was also positively correlated to the malate signal at 4.28 ppm. Metal taste did not show any correlations above $|R| = 0.5$ with other sensory descriptors in this study.

Very surprisingly, redness was positively correlated with the presence of tryptophan (H4, H5, and H6 at 7.75, 7.17, and 7.29 ppm, respectively) and tyrosine (H3/H5 at 6.90 and H- β s at 3.06 and 3.18 ppm). At this stage, we cannot explain this observation.

Finally, a number of signals in the region between 4.30 and 4.60 ppm and at 4.03 ppm display negative correlations with density. For the time being, we are not able to unambiguously assign these signals, even if their chemical shifts strongly suggest that they could be attributed to sugars.

In conclusion, the perception of odor and flavor of food is a complicated physiological and psychological process that cannot be explained by simple models. This is because hundreds of compounds simultaneously influence the human olfactory receptors and because the physiological response is far from linear, and the overall effects are not just the superimposition of the effect of single stimuli.

Sensory analysis, and, in particular, the QDA, continues to be an irreplaceable technique to describe sensory features. Nevertheless, the availability of a number of instrumental techniques has opened up the possibility to calibrate the sensory perception. Thus, the tandem approach that uses instrumental and classical sensory analysis seems to be a valuable strategy. Unfortunately, the more usual artificial tongue/nose are used to determine very specific components of the analyzed food. Furthermore, not all instrumental techniques are able to analyze directly the genuine mixture interacting with our sense without any extraction/concentration procedures. For example, MS and GC require volatilization of the analyzed compounds that very often is obtained with a chemical derivatization. In this frame, we have tried to test the potentiality of NMR spectroscopy as a predictive tool to measure sensory descriptors, without performing any complementary chemical analyses. In particular, we have used an NMR

metabolomic approach since it is rapid, sensitive, and relatively inexpensive. This approach in combination with multivariate analysis has an advantage over the ordinary sensory test, since it offers more reliable results for the classification and determination of some aspect of the sensory attribute of the tomato. The metabolomic fingerprints recorded for all tested canned tomato samples allowed us to differentiate all analyzed samples based on their chemical composition.

Interestingly, the same classification and characterization have been reached independently from the QDA analysis. In particular, a number of sensory descriptors can be easily predicted from the NMR data: bitterness, sweetness, sourness, saltiness, tomato and metal taste, redness, and density. The presence of a number of bitter amino acids like isoleucine, tryptophan, tyrosine, phenylalanine, and valine is correlated with bitterness and surprisingly to sourness. Other amino acids seem also to have a crucial role as taste enhancers like glutamate, glutamine, aspartate, and asparagine, which amplify the bitter and the sour taste, as well as the cooked taste. The sugar content is obviously correlated with sweetness, even if their correlation is not so straightforward. Finally, other components like citrate, malate, formate, and acetate are correlated with sourness. Very interestingly, citrate and particularly malate seem to be crucial in the defining the taste of tomato. In general, we have noted that the same substances could be involved in two (or more) features; these could be counteractive in the sense that the increase in one leaves less room for the other features; they could be also affected by a third feature, etc. One drawback with the methodology presented here is that only the soluble fraction of the product is measured. In future studies, this can be avoided by using HR-MAS NMR where also the semisolid fractions contribute to the NMR spectrum. However, the results obtained suggest that NMR could be a very useful tool for the characterization of some sensory features of tomato. To evaluate the applicability of this methodology to other kinds of food, a number of experiments are currently undertaken in our laboratories.

■ ASSOCIATED CONTENT

S Supporting Information. Tables of the identities of the canned tomatoes used in this study (Table S1) and of the QDA mean results forming the basis for Figure 1 (Table S2). This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Authors

*Tel: +45-3532-7753. Fax: +45-3535-6310. E-mail: malmendal@sund.ku.dk (A.M.). Tel: +39-081-678514. Fax: +39-081-678552. E-mail: antonio.randazzo@unina.it (A.R.).

Funding Sources

A.M. acknowledges support from the Danish Research Foundation.

■ ACKNOWLEDGMENT

We are grateful to Anna Mannara for her precious work.

■ REFERENCES

(1) (a) Noble, A. C. Sensory methods of flavor analysis. In *Food Flavor Technology*; Taylor, A. J., Ed.; Sheffield Academic Press: Sheffield, United Kingdom, 2002; pp 252–275. (b) Stone, H.; Sidel,

J. L. Quantitative descriptive analysis: Developments, applications, and the future. *Food Technol.* **1998**, *52* (8), 48–52.

(2) Deising, A. K.; Stone, D. C.; Thompson, M. Application of electronic noses and tongues in food analysis. *Int. J. Food Sci. Technol.* **2004**, *6*, 587–604.

(3) Bard, A. J.; Faulkner, L. R. *Electrochemical Methods—Fundamentals and Applications*; John Wiley & Sons Inc.: New York, NY, 1980.

(4) Legin, A.; Rudnitskaya, A.; Vlasov, Y.; Di Natale, C.; Davide, F.; D'Amico, A. Tasting of beverages using an electronic tongue based on potentiometric sensor array. *Technical Digest of Eurosensors X*; Leuven, Belgium, 1996; pp 427–430.

(5) Winquist, F.; Wide, P.; Lundström, I. An electronic tongue based on voltammetry. *Anal. Chim. Acta* **1997**, *357*, 21–23.

(6) Lavigne, J. J.; Savoy, S.; Clevenger, M. B.; Ritchie, J. E.; McDoniel, B. S.; Yoo, J.; Anslin, E. V.; McDevitt, J. T.; Shear, J. B.; Neikirk, D. Solution-based analysis of multiple analytes by a sensor array: Toward the development of an "Electronic Tongue". *J. Am. Chem. Soc.* **1998**, *120*, 6249–6430.

(7) (a) Ballantine, D. S.; White, R. M.; Martin, S. J.; Ricco, A. J.; Zellers, E. T.; Frye, G. C.; Wohltjen, H. *Acoustic Wave Sensors Theory, Design and Physico-Chemical Applications*; Academic Press: San Diego, CA, 1997. (b) Herrmann, F.; Jakoby, B.; Rabe, J.; Buttgenbach, S. Microacoustic Sensors for Liquid Monitoring. In *Sensors Update*; Baltes, H.; Hesse, J.; Korvink, J., Eds.; Wiley-VCH: Weinheim, Germany, 2001; No. 9, pp 105–160.

(8) (a) Aishima, T. Correlating sensory attributes to gas chromatography-mass spectrometry profiles and e-nose responses using partial least squares regression analysis. *J. Chromatogr. A* **2004**, *1054*, 39–46. (b) Zanor, M. I.; Rambla, J.; Chaib, J.; Steppa, A.; Medina, A.; Granell, A.; Fernie, A. R.; Causse, M. Metabolic characterization of loci affecting sensory attributes in tomato allows an assessment of the influence of the levels of primary metabolites and volatile organic contents. *J. Exp. Bot.* **2009**, *60*, 2139–2154.

(9) Taylor, A. J.; Rob, S. T.; Linforth, R. S. T. Direct mass spectrometry of complex volatile and non-volatile flavour mixture. *Int. J. Mass Spectrom.* **2003**, *223*, 223–224, 179–191.

(10) (a) Rochfort, S.; Ezeriye, V.; Bastian, S. E. P.; Downey, M. O. Sensory attributes of wine influenced by variety and berry shading discriminated by NMR metabolomics. *Food Chem.* **2010**, *121*, 1296–1304. (b) Skogerson, K.; Runnebaum, R.; Wohlgemuth, G.; de Ropp, J.; Heymann, H.; Fiehn, O. Comparison of Gas Chromatography-Coupled Time-of-Flight Mass Spectrometry and H-1 Nuclear Magnetic Resonance Spectroscopy Metabolite Identification in White Wines from a Sensory Study Investigating Wine Body. *J. Agric. Food Chem.* **2009**, *57*, 6899–6907.

(11) Tomasi, G.; van den Berg, F.; Andersson, C. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *J. Chemom.* **2004**, *18*, 231–241.

(12) Keun, H. C.; Ebbels, T. M. D.; Antti, H.; Bollard, M. E.; Beckonert, O.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Improved analysis of multivariate data by variable stability scaling: Application to NMR-based metabolic profiling. *Anal. Chim. Acta* **2003**, *490*, 265–276.

(13) Trygg, J.; Wold, S. Orthogonal projections to latent structures, O-PLS. *J. Chemom.* **2002**, *16*, 119–128.

(14) Sobolev, A. P.; Segre, A.; Lamanna, R. Proton high-field NMR study of tomato juice. *Magn. Reson. Chem.* **2003**, *41*, 237–245.

(15) Petro-Turza, M. Flavor of tomato and tomato products. *Food Rev. Int.* **1986**, *2* (3), 309–351.

(16) deBruyn, J. W.; Garretsen, F.; Kooistra, E. Variation in taste and chemical composition of the tomato (*Lycopersicon esculentum* Mill.). *Euphytica* **1971**, *20*, 214–227.

(17) Davies, J. N.; Kempton, R. J. Changes in the individual sugars of tomato fruit during ripening. *J. Sci. Food Agric.* **1975**, *26*, 1103–1110.

(18) Jones, R. A.; Scott, S. J. Genetic potential to improve tomato flavour in commercial F₁ hybrids. *J. Am. Soc. Hortic. Sci.* **1984**, *109*, 318–321.

(19) Watada, A. E.; Aulenbach, B. B. Chemical and sensory qualities of fresh market tomatoes. *J. Food Sci.* **1979**, *44*, 1013–1016.

(20) Stevens, M. A.; Kader, A. A.; Albright-Holton, M.; Algazi, M. Genotypic variation for flavor and composition in fresh market tomatoes. *J. Am. Soc. Hortic. Sci.* **1977**, *102*, 680–689.

- (21) Yoshida, M.; Saito, S. Multidimensional scaling of the taste of amino acids. *Jpn. Psychol. Res.* **1969**, *11*, 149–166.
- (22) Amerine, M. A.; Pangborn, R. M.; Roessler, E. B. *Principles of Sensory Evaluation of Food*; Academic Press: New York and London, 1965.
- (23) Kirimura, J.; Shimizu, A.; Kimizuka, A.; Ninomiya, T.; Katsuya, N. The contributi peptides and amino acids to the taste of foodstuff. *J. Agric. Food Chem.* **1969**, *17*, 689–695.
- (24) Solms, J. The taste of amino acids, peptides, and proteins. *J. Agric. Food Chem.* **1969**, *17*, 686–688.
- (25) Davies, J. N. Effect of nitrogen, phosphorus and potassium fertilisers on the non-volatile organic acids of tomato fruit. *J. Sci. Food Agric.* **1964**, *15*, 665–672.
- (26) Mahakum, N.; Leeper, P. W.; Burns, E. E. Acidic Constituents of various tomato fruit types. *J. Food Sci.* **1979**, *44*, 1241–1244.
- (27) Luh, B. S.; Daoud, H. N. Pectin, amino acids and carotenoids in tomato juices. *Fruchtsaft-Ind.* **1968**, *13*, 204.

Supplementary Material

Table S1. Identities of the canned tomatoes used in this study.

Sample #	Sample Brand
1	Valfrutta - Gran Cubetti di Giornata
2	De Rica - Polpa Pronta
3	Primadonna - Polpa di Pomodoro
4	Annalisa - Polpa di Pomodoro
5	Santa Rosa Bertolli - I Pezzettoni
6	Esselunga - Polpa di Pomodoro
7	Carrefour - Polpa di Pomodoro
8	Feger - Polpa di Pomodoro
9	Coop - Polpa di Pomodoro
10	Delizia del Sole - Polpa di Pomodoro a Pezzetti
11	Mutti - Polpa
12	Tesori dell' Arca - Polpa di Pomodoro
13	Cirio - PolpaPiu'
14	Cirio - Polpadoro
15	Cirio - Fior di Filetti
16	Star - Polpabella
17	De Rica - Polpa di Pomodoro
18	La Doria - Polpa di Pomodoro

Table S2. QDA mean results.

	Redness	Dimension	Residual peel	Synaeresis	Density	Tomato flavor	Saltiness	Sourness	Sweetness	Bitterness	Tomato taste	Cooked taste	Metal taste	Consistency
1	5.5	5.3	6.4	4.7	6.5	3.5	3.1	4.8	4.4	2.6	4.3	6.0	4.9	3.4
2	5.5	4.9	3.8	5.4	6.3	5.5	6.8	2.8	9.0	0.7	6.4	4.5	2.1	3.8
3	3.4	3.5	4.1	7.1	2.0	3.5	2.7	6.1	1.8	3.2	4.9	4.0	4.3	3.2
4	2.3	5.2	4.8	5.9	4.0	1.3	1.1	3.1	3.0	2.9	1.4	3.0	4.4	3.4
5	5.9	7.0	5.7	6.8	5.5	4.3	8.6	5.7	4.8	1.7	5.3	8.2	4.2	6.0
6	5.4	3.7	5.6	4.0	6.3	4.3	2.2	3.9	3.7	2.7	4.5	3.3	4.6	3.5
7	3.7	4.5	5.4	6.5	3.7	3.5	5.0	1.5	5.1	2.1	4.3	7.0	4.2	3.4
8	5.3	6.3	4.0	7.1	5.9	5.5	5.6	3.4	5.2	1.0	6.3	5.5	2.8	5.9
9	6.5	6.4	4.0	6.8	6.0	5.6	2.3	4.7	3.5	4.7	5.2	4.3	4.9	6.0
10	6.5	5.4	3.3	5.5	5.3	4.6	3.0	8.3	3.8	3.6	4.4	5.8	5.5	5.0
11	3.3	0.1	0.4	5.5	0.9	1.3	1.1	2.8	4.2	2.5	1.6	3.3	5.5	0.1
12	6.3	5.3	4.6	3.5	6.7	3.6	6.0	3.4	4.7	2.3	5.5	5.7	3.7	4.7
13	6.7	5.0	7.1	4.3	8.8	8.5	6.3	3.2	5.4	1.5	6.6	5.5	4.5	3.8
14	5.2	3.1	1.4	2.3	7.4	6.8	6.7	3.5	6.1	0.8	7.2	4.7	2.8	4.5
15	8.6	9.9	1.8	2.2	7.4	4.9	2.8	5.2	4.0	3.8	5.2	6.3	4.6	6.4
16	9.1	5.8	5.7	4.6	6.9	4.0	5.0	5.3	4.5	4.4	5.0	5.8	5.3	4.7
17	4.9	5.4	5.5	3.1	6.4	4.3	3.7	3.1	5.2	1.8	5.6	3.7	8.3	4.5
18	8.7	5.7	5.7	3.6	9.0	4.7	9.0	5.1	4.6	2.7	4.0	8.0	4.5	4.0
LSD	0.7	0.5	0.7	0.8	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8

LSD 95% = Two products are significantly different for each attribute when the difference of mean values are \geq LSD value.

Paper II

Lauri I., Pagano B., Malmendal A., Sacchi R., Novellino E.,
Randazzo A.

Application of “magnetic tongue” to the sensory
evaluation of extra virgin olive oil

Food Chemistry, 2013, 140, 692–699.



Application of “magnetic tongue” to the sensory evaluation of extra virgin olive oil

Ilaria Lauri^a, Bruno Pagano^a, Anders Malmendal^b, Raffaele Sacchi^c, Ettore Novellino^a, Antonio Randazzo^{a,*}

^a Dipartimento di Chimica Farmaceutica e Tossicologica, Università degli Studi di Napoli “Federico II”, Via D. Montesano 49, 80131 Napoli, Italy

^b Department of Biomedical Sciences, Faculty of Health Sciences, University of Copenhagen, DK-2200 Copenhagen N, Denmark

^c Dipartimento di Scienza degli Alimenti, Università degli Studi di Napoli “Federico II”, Via Università 100, 80055 Portici, Napoli, Italy

ARTICLE INFO

Article history:

Available online 10 November 2012

Keywords:

NMR

Multivariate analysis

Sensory analysis

Extra virgin olive oil

ABSTRACT

The perception of odour and flavour of foods is a complicated physiological and psychological process that cannot be explained by simple models. Unfortunately, taste is not objective, but partially subjective and it depends also on the mood of the taster. Generally, sensory analysis is used to describe sensory features. The availability of a number of instrumental techniques has opened up the possibility to calibrate the sensory perception. Here we have tested the potentiality of nuclear magnetic resonance spectroscopy as “magnetic tongue” to measure sensory descriptors in extra-virgin olive oil. We were able to correlate the NMR metabolomic fingerprints of extra-virgin olive oil to the sensory descriptors: tomato, bitter, pungent, rosemary, artichoke, sweet, grassy and leaf.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

The sensory impression of a food is determined mainly by the chemical senses of taste and smell. They are both sensed through sensory cells of the tongue (taste) and of the nasal cavity (smell) (Roper, 2006). Sensory cells are able to differentiate between the different tastes and smells based on different molecules or ions.

From a food company point of view, understanding how much consumers' preferences are driven by differences in sensory features between products is extremely important. Traditional consumer research helps determining acceptable versus unacceptable. It is helpful when an overall, synthetic understanding of the products acceptance is needed. However, it is not of any help when an explanation, in terms of sensory descriptors, is needed in order to provide R&D with technical information useful to enhance product features. Such information can only be provided through analytical products evaluation, of which consumers are not capable. For these reasons, food and beverages industries take advantage of quantitative descriptive analysis (QDA) (Noble, 2002; Stone & Sidel, 1998). This is a discipline through which the sensory analyst evokes, measures, analyzes and interprets human responses to stimuli as perceived through the senses. Unfortunately, taste is not objective like for example the sight, but partially subjective and it also depends also on the mood of the taster. Thus a number of more objective analytical techniques have been used to support or, in some cases, replace the classical QDA. Among these, it is worthy to mention the electronic nose and the electronic tongue (Deising,

Stone, & Thompson, 2004). Furthermore, techniques like mass spectrometry (MS) (Aishima, 2004; Zanol et al., 2009) and gas chromatography (GC) (Taylor, Rob, & Linforth, 2003) have also been used. Very recently, we have proposed the use of the ¹H nuclear magnetic resonance (NMR) spectroscopy as “magnetic tongue” to predict the sensory descriptors of canned tomatoes (Malmendal et al., 2011). Herein, we intend to widen the applicability of this technique testing the ¹H-NMR spectroscopy to predict the taste of extra virgin olive oil (EVOO).

2. Materials and methods

2.1. Materials

Eighteen EVOO products of different brands were given from different companies in the Campania region (Italy) within the EXTRABIO 2008, a quality prize organised by the Chamber of Commerce of Naples for olive oils from organic agriculture.

2.2. Sensory assessment

Sensory profiles of the 18 samples were determined by the olive oil sensory panel of the “Laboratorio Chimico Merceologico” of the Chamber of Commerce of Naples (Italy). This panel is particularly trained to recognise, describe and quantify basic taste and odour properties. Evaluation took place in individual testing booths according to the official method (EC Regulation 2568/91). Eleven descriptors have been defined: fruity, leaf, grassy, bitter, pungent, sweet, almond, artichoke, apple, tomato and rosemary tastes.

* Corresponding author. Tel.: +39 081 678514; fax: +39 081 678552.

E-mail address: antonio.randazzo@unina.it (A. Randazzo).

Descriptors were evaluated on a continuous, unlabeled, 0–10 intensity scale, and then turned into numeric variables (a number between 0 and 10). Oils were served in coloured tasting glasses. The temperature of the oils was kept constant (28 ± 2 °C). Samples were labelled with a 3-digit code and served based on a balanced rotation plan.

2.3. Sample preparation for ^1H -NMR analysis

The procedure previously reported by Segre and Mannina (1997) was followed to prepare the samples: 20 μL EVOO were dissolved in 560 μL chloroform- d and 20 μL DMSO- d in 5 mm NMR tubes.

2.4. NMR spectrometry

^1H -NMR spectra were acquired at 25 °C with a 700 MHz Varian Unity Inova spectrometer using a 5 mm $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ triple resonance probe. The ^1H -NMR measurements were carried out with 1000 transients and 32 K complex data point. In order to retrieve quantitative information, the recycle time was set to 5 s, and a 45° pulse angle was used.

2.5. NMR data reduction and processing

The spectra were processed using iNMR (www.inmr.net). An exponential line-broadening of 0.5 Hz was applied to the free-induction decay prior to Fourier transformation. All spectra were referenced relative to external sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS), phased and baseline corrected. The spectra were aligned by using the Icoshift algorithm (Savorani, Tomasi, & Engelsen, 2010). Data reduction was accomplished by dividing the spectrum into 0.005 ppm regions (bins) over which the signal was integrated to obtain the signal intensity. Only the spectral region between 6.55 and 9.75 (excluding the region around the chloroform signal between 7.0 and 7.8 ppm) was considered for the study (see Section 3). At the end, a total of 480 variables were analysed for each spectrum. The integrals were normalised to the integral of the triplet at δ_{H} 0.86 ppm (CH_3 of triacylglycerols) in order to suppress trivial separation based on variations in the amount of sample.

2.6. Statistical analysis

Statistical analysis was performed using Simca-P 13.0 (Umetrics, Umeå, Sweden). The number of principal components (PCs) in the principal component analysis (PCA) (Eriksson, Johansson, Kettaneh-Wold, & Wold, 2006) was determined by leave one out cross-validation, where all the samples from each of the 18 products were left out for one product at a time to determine the quality of the model. The model used was estimated using all 18 products. Sensory data were Unit-Variance (UV) scaled, while NMR data were pareto-scaled (Eriksson, Johansson, Kettaneh-Wold, & Wold, 1999), where each value has been divided by the standard deviation and by the square root of the standard deviation computed around the mean, respectively. Both dataset were also centred. Two PCs were computed for both sensory and NMR PCAs.

The dendrograms for the sensory analysis and NMR data were both based on the PCA scores (PC1 and PC2). Hierarchical cluster analysis (HCA) was carried out by using Ward clustering method (Ward, 1963).

Orthogonal projection to latent structures, OPLS, separates the variance in x correlated with y (y -predictive) with the orthogonal (non-correlated; y -orthogonal) variance (Trygg & Wold, 2002). In contrast to regular PLS, a single y will result in only one predictive

component. OPLS was carried out using each sensory descriptor as the y -variable. As in PCA, sensory data were UV-scaled, while NMR data were pareto-scaled and both centred. Cross validation was obtained as described above. Markers for the sensory descriptors were identified from the NMR signals that showed a strong correlation ($R^2 > 0.5$) with the OPLS predictive scores for the sensory descriptors.

3. Results and discussion

3.1. Sensory analysis

Eighteen samples of EVOO were selected from local industries of the Campania region in Italy. These samples have been judged by assessors specifically trained in the application of quantitative descriptive analysis (QDA) and in the evaluation of EVOO.

The QDA method aims to (i) define a product's sensory profile (describing products in terms of sensory features as perceivable through five senses), (ii) identify similar products sharing similar sensory properties through cluster analysis, (iii) define a sensory map of products through principal component analysis (PCA), in order to position, within an overall picture, products in terms of their sensory properties. Main sensory descriptors were identified and selected (as actually describing products and discriminating among them) during several preliminary sessions. These are: fruitiness, leaf, grassy, bitter, pungent, sweet, almond, artichoke, apple, tomato and rosemary tastes.

During the sessions of analysis, assessors determined the intensity of each attribute for each product through objective evaluations. Evaluations were given on a continuous, no labelled, 0–10 scale, and then turned into numeric variables (a number between 0 and 10). Products were tested plain and not cooked. QDA mean results are reported in Fig. 1 (Table S1 in Supplementary material). In order to group products sharing similar sensory features, hierarchical cluster analysis (HCA) was performed on QDA means. The resulting dendrogram is shown in Fig. 2A. Basically 3 groups have been identified: samples 13, 4, 9, 5, 8, 2, 3, 14 (Group 1A); samples 18, 12, 1, 7 (Group 2A); samples 6, 15, 10, 17, 11, 16 (Group 3A).

Principal component analysis (PCA) was also performed on the same dataset (Fig. 3). Two principal components (PCs) accounting for 65% of the variation were identified. The plot of their scores (Fig. 3A) shows the positioning of the products according to their sensory attributes and allowed the identification of the most important sensory descriptors for products differentiation. Interestingly, a number of descriptors shows a significant correlation ($Q^2(\text{cum}) > 0.4$) with the model (Table S2), indicating a high level of correlation between the sensory descriptors. Almond taste is not predicted by the model ($Q^2(\text{cum}) = -0.07$). PCA indicates that the groups identified by the HCA share the same features, and that there is no strong separation between the different groups identified.

According to the loading plot (Fig. 3B), transition from the left to the right of the map shows the simultaneous decrease of the sweet taste and increase of the bitter, pungent, fruity and artichoke tastes. The sensory map is also influenced by apple, tomato, grassy tastes that stand on the top of the map, opposite to leaf and rosemary tastes (on the bottom). In general, Group 1A, which is the most populated group, contains products that are characterised by apple, tomato, grassy, artichoke, pungent, fruity and bitter tastes. On the other hand, Group 3A contains products with more marked sweet taste. Group 2A, instead, is characterised mostly by rosemary and leaf tastes.

The loadings plot (Fig. 3B) suggests a covariance between different tastes. Thus a higher sweetness is generally associated with a lower bitterness. It seems there is also a strong correlation

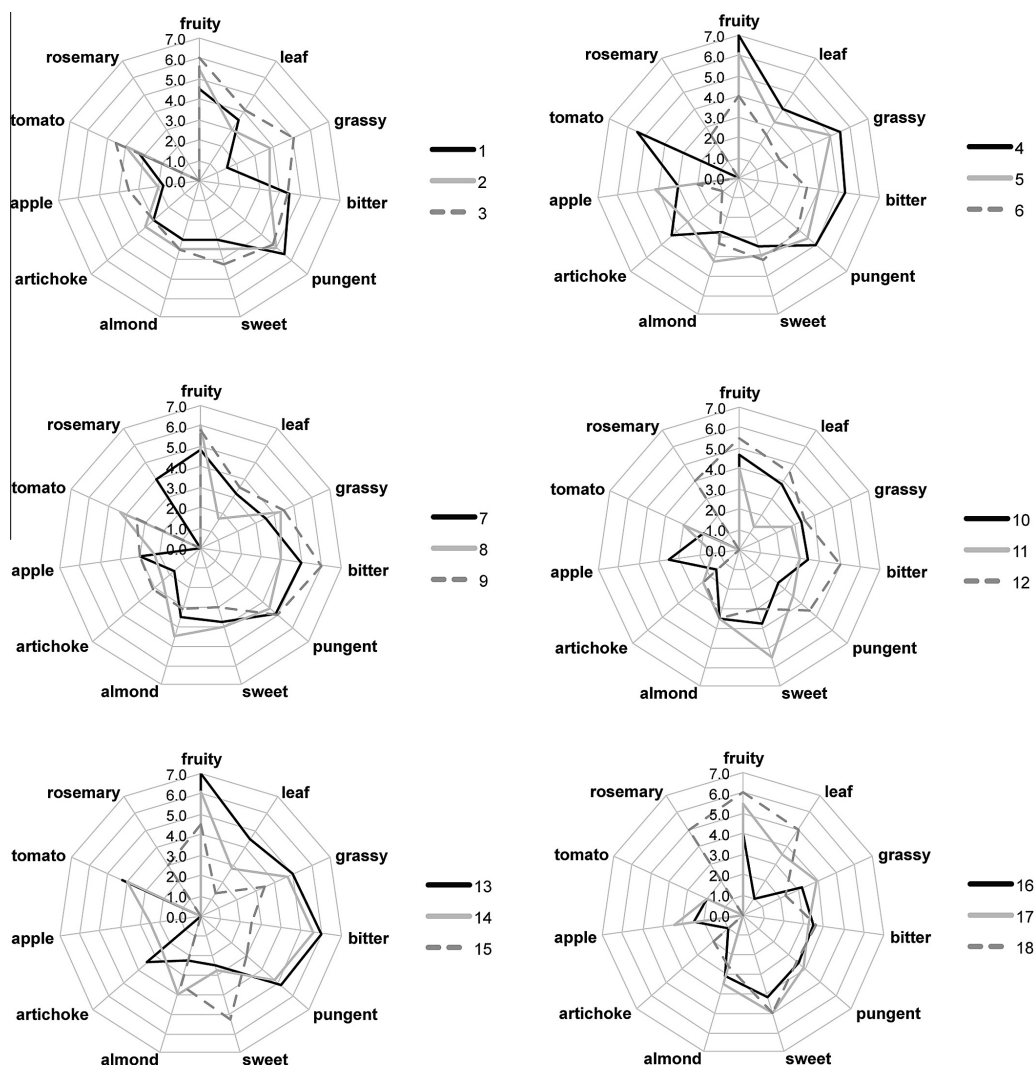


Fig. 1. Spider-web plot of the sensory descriptors for the 18 tested samples. The mean QDA parameters are listed in Table S1 in the Supplementary information. In order to group products sharing similar sensory features, hierarchical cluster analysis (HCA) has been performed on QDA means. The resulting dendrogram is shown in Fig. 2A.

between artichoke, pungent, fruity and bitter tastes and so forth. To further characterise the correlations between different sensory descriptors, the correlation coefficients were calculated (Table 1). Interestingly, there is a strong positive correlation between bitter, pungent and artichoke tastes. These are, in turn, negatively correlated to sweet taste. There is also a strong correlation between grassy and fruity tastes. As expected, from the PCA analysis, no significant correlation could be calculated for almond taste.

3.2. NMR analysis

The same products tested by QDA were analysed by NMR. The rationale is to use an NMR metabolomic approach to compare

the constituent of the different EVOO samples, and correlate the NMR data with the sensory descriptors obtained by the QDA analysis.

Fig. 4 shows different vertical expansions of the NMR spectrum of a representative sample of EVOO. Particularly, Fig. 4A shows the NMR spectrum at a regular scale, where the major signals of aliphatic and glyceryl protons of triacylglycerols, including allylic and diallylic resonances of monounsaturated and polyunsaturated fatty acids, can be easily singled out.

Increasing the vertical scale (Fig. 4B) also diacylglycerols (sn-1,2 and sn-1,3), naturally present in EVO oil at level of about 2–4% (as molar fraction), methyls of linolenic acid (18:3 *n*–3) present at levels less than 1% can be detected.

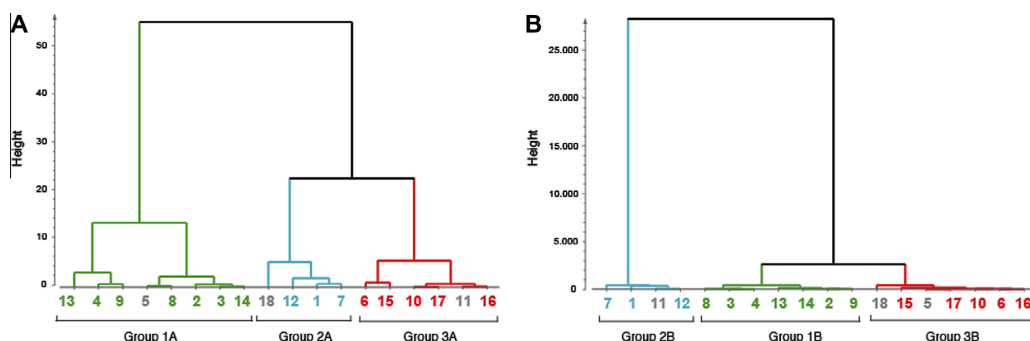


Fig. 2. Dendrograms showing similarities between products based on QDA (A) and NMR (B). The dendrograms were based on the two first PCs after PCA of QDA (A) and NMR (B) data, respectively. Products falling within the same group are indicated with the same colour. Most of the products are grouped in the same way using both QDA and NMR (Group 1A/1B; Group 2A/2B; Group 3A/3B). Products that are not grouped in the same way are coloured in grey. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

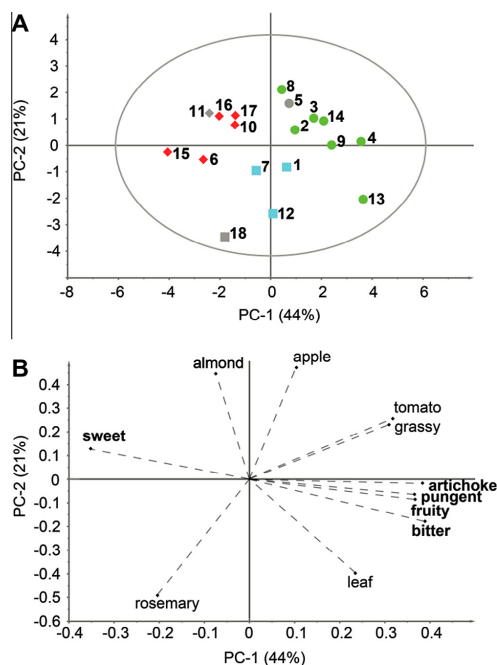


Fig. 3. Score (A) and loading (B) plots of the PCA performed on sensory data. Products are coloured according to QDA HCA analysis in Fig. 2. Sample belonging to Group 1A, 2A and 3A have circle, box and diamond signs, respectively. Variables in loading plot (B) that are well predicted by the model ($Q^2(\text{cum}) > 0.4$) are reported in bold. Almond, that cannot be predicted and hence has a small impact on the model, is reported in light grey. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Human sensory cells cannot detect the molecules described so far and all the EVOO have basically identical composition of these constituents. Hence, the sensory perception of the EVOO is related to very minor components. These can be observed by further

increasing the vertical scale of the spectrum (Fig. 4C), which allows the identification of resonances of very minor components like, for example, different aldehyde CHO protons (*n*-alcanals, *n*-alchenals, idroxy-alchenals, seicoridoids aglycons) (Sacchi, Addeo, & Paolillo, 1997) in the region around 9 ppm and phenolic ring signals around 6–7 ppm (Christophoridou & Dais, 2009). These last molecules are commonly attributed the bitter-pungent taste of EVOO, with a main role of phenol compounds (Andrewes et al., 2003) in combination to selected volatiles (Caporale, Policastro, & Monteleone, 2004).

Thus, in order to study EVOO from taste point of view, we had to consider only the spectral regions that were not occupied by the signals of the taste-free major constituents. Furthermore, it should also be noted that almost all the analysed samples contain small amounts of water from the production of the EVOO. The signal of water in the NMR spectra of olive oil acquired in chloroform and dimethyl sulfoxide (see Section 2) is very broad and sometime very intense, and generally resonates between 4 and 6 ppm. This means that this range cannot be used to compare different EVOO samples either. Furthermore, the spectral regions containing solvent signals cannot be taken into account. Thus, in our case, the regions that can be considered are from 6.55 ppm to 7.00 and from 7.80 to 9.75 ppm. Other smaller regions could also be considered, but they are dramatically affected by baseline distortion from the very intense triacylglycerol signals that offset the intensity values, thus resulting in inaccuracy in peak assignment and quantification.

Analogously to the HCA performed on sensory data, the HCA analysis performed on the NMR data revealed three main groups (Fig. 2B): 8, 3, 4, 13, 14, 2, 9 (Group 1B); 7, 1, 11, 12 (Group 2B); 18, 15, 5, 17, 10, 6, 16 (Group 3B). Despite the fact that the two HCAs refer to data collected by very different analytical techniques, it can be seen that there is an excellent global agreement between the different measurements: only sample 5, 11 and 18 are not grouped in the same way using QDA and NMR data.

PCA has also been performed on the NMR dataset. Two PCs were identified, accounting for 83.6% of the variation. The general distribution of the products in the score plots (Fig. 5A) in a way recall the one observed in the sensory dataset (Fig. 3A). For example, Group 1B and Group 3B are mapped opposite to each other, suggesting that PC2 in the PCA computed with the NMR data is in a way related to the sweetness/bitterness of the samples. On the other hand Group 2B is very well separated from the other two groups along PC1, suggesting that this PC is related to apple, rosemary and leaf tastes.

Table 1
Correlation coefficients ($|R| > 0.5$) between sensory descriptors.^a

	Fruity	Leaf	Grassy	Bitter	Pungent	Sweet	Artichoke	Apple	Tomato	Rosemary
Fruity		0.62	0.77	0.63	0.51		0.63			
Leaf	0.62			0.50						
Grassy	0.77									
Bitter	0.63	0.50			0.75 ^b	−0.84	0.63			
Pungent	0.51			0.75		−0.70	0.71			
Sweet				−0.84	−0.70		−0.60			
Artichoke	0.63			0.63	0.71	−0.60			0.61	
Apple							0.61			−0.54
Tomato								−0.54		−0.75
Rosemary									−0.75	

^a Sensory descriptors showing $|R| < 0.5$ to all other sensory descriptors are excluded.

^b The correlation coefficients between bitter, pungent, sweet and artichoke tastes are italics.

As judged from the loading plots (Fig. 5B and C), Group 2B contains higher concentration of molecules having signals at δ_H 9.58, 9.31, 7.88, 6.97, 6.69 and 6.63 ppm, and low concentration of molecules having signals at δ_H 9.45, 9.09, 9.05, 6.82 and 6.58 ppm, relative to Groups 1B and 3B.

However, Groups 1B and 3B are differentiated along PC2. Looking at the loading plot of second principal component (Fig. 5C), it is clear that Group 1B contains higher concentrations of molecules

having signals at δ_H 9.45, 9.09, 9.05, 6.82, 6.58, whereas lower concentrations of those molecules are present in the Group 3B.

3.3. Prediction of sensory descriptors

In order to determine how well the sensory descriptors can be predicted by NMR, we made predictive models for the different

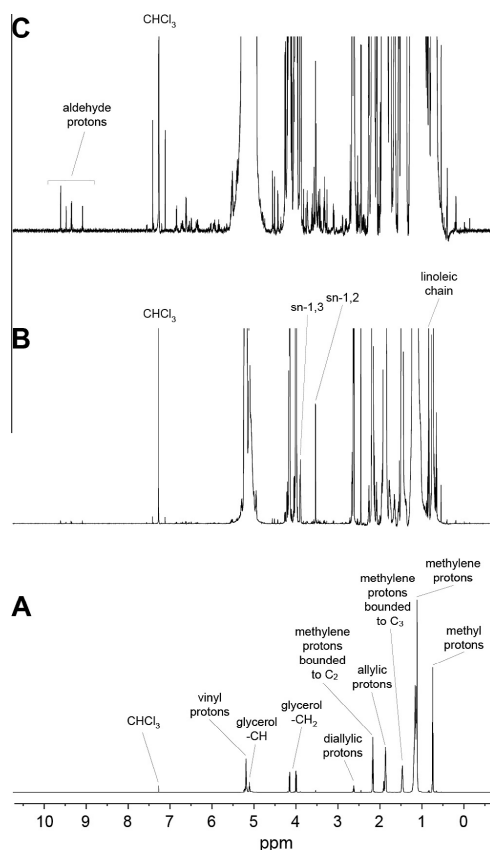


Fig. 4. Increasing (from A to C) vertical expansions of the NMR spectrum of a representative sample of extra-virgin olive oil.

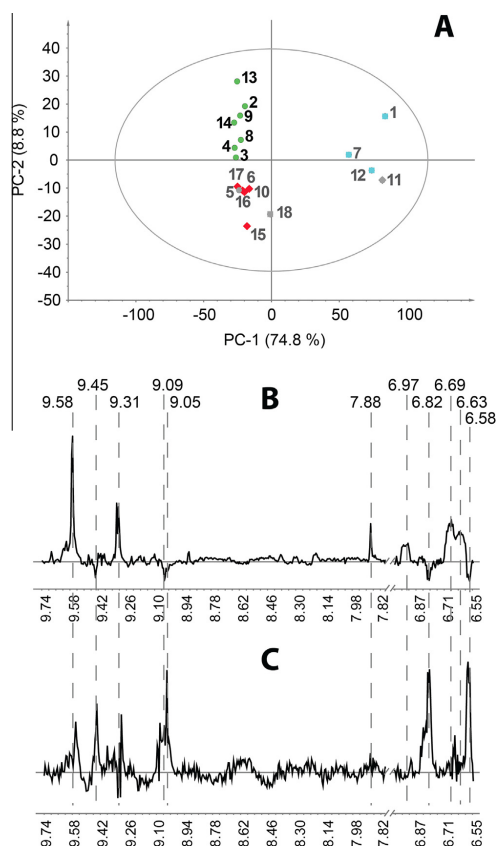


Fig. 5. Score (A) and loading (B and C) plots of the PCA performed on NMR data. (B and C) show the PC1 and PC2 loadings. Products are coloured according to Fig. 2.

sensory descriptors using orthogonal-projection to latent structures, OPLS (Trygg & Wold, 2002). Using this protocol we were able to get good predictions ($Q^2(\text{cum}) > 0.4$) for tomato, bitter, pungent rosemary and artichoke tastes (Table 3). After further inspection of all models, and identification of outliers we were able to get good models for all but 1, 2, 3 and 4 brands for sweet, grassy, artichoke and leaf taste, respectively (Table 2). In seven cases out of ten the removed products showed extreme values for sweet (1 of 1), grassy (2 of 2), artichoke (2 of 3) and leaf (2 of 4) tastes. It thus seems that these extra strong features depend on other factors than under more normal conditions.

3.4. Chemical signatures of sensory descriptors

In order to determine which chemical components are responsible for a given sensory descriptor, we have looked for all possible correlations between the NMR signals (in the considered spectral regions) and the analysed sensory descriptors using OPLS models. In order to have a qualitative view of the correlations, we first had a look at the loading plots of the predictive component of each OPLS model having with $Q^2 > 0.4$ (Fig. 6).

Bitter, pungent and artichoke tastes are highly correlated ($R > 0.63$) and display a very similar profile. They all show a strong

Table 2

Description and statistical summary of the OPLS models constructed based on NMR data.^a

Variable	A ^b	N ^c	R ² X(cum) ^d	R ² Y(cum) ^d	Q ² (cum) ^e
Tomato	2	18	0.86	0.95	0.82 ^a
Bitter	4	18	0.91	0.98	0.65 ^a
Pungent	1	18	0.84	0.76	0.52 ^a
Rosemary	5	18	0.92	0.99	0.43 ^a
Artichoke	1	18	0.83	0.56	0.42
Artichoke excluding products 4, 5 and 17	2	15	0.86	0.93	0.71 ^a
Sweet	1	18	0.83	0.73	0.39 ^a
Sweet excluding product 11	1	17	0.82	0.80	0.72 ^a
Grassy	2	18	0.87	0.82	0.39 ^a
Product 1 and 5 excluded	9	16	0.96	1.00	0.72 ^a
Fruity	1	18	0.83	0.55	0.33 ^a
Leaf excluding products 11, 12, 17 and 18 ^f	3	14	0.88	0.93	0.53

^a No significant OPLS model could be constructed for almond and apple tastes, that are excluded from the table.

^b A number of orthogonal components.

^c Number of samples included in the model.

^d R²X(cum) and R²Y(cum) = The cumulated fraction of the variance in the parameter explained by the model.

^e Q²Y(cum) = the cumulative predicted fraction of the variation of the parameter as determined by cross-validation. Q²Y(cum) values above 0.4 are considered as good predictors.

^f Leaf taste can be well predicted (Q²(cum) > 0.5) only excluding product 11, 12, 17 and 18.

Table 3

Correlation among NMR signals and OPLS models for sensory descriptors.^a

Ppm (δ_H)	Bitter	Pungent	Artichoke	Sweet	Rosemary	Tomato	Grassy	Fruity	Leaf
9.70 (hexenal)				—	+	—	—	—	+
9.58[*] (4-hydroxy- <i>trans</i> -alk-2-enal) ^{**}	+	+	+	—	+	—	—	—	+
9.55 (alk-2,4-dienals) ^{**}							—		
9.47						—	—	—	
9.45 (<i>trans</i> -2-hexenal)						+	+	+	
9.39	+	+	+	—	—	+	+		
9.31		+	+	—	+	—	—	—	
9.25						+	+		+
9.21		—			+	—	—		
9.11					+	—			
9.09 (secoiridoids)	+	+	+	—	+	+			
9.05	+	+	+	—	—	+	+	+	+
8.95							+		
8.91							—	—	
8.76					+	—	—	—	
8.61					—	+	+	+	
8.47					+	—	—	—	
8.19				—	+	—	—	—	
7.88				—	+	—	—	—	+
6.94					+	—	—	—	+
6.84 (<i>trans</i> -alk-2-enals) ^{**}	+	+	+	—	—	+	+	+	+
6.80	+	+	+	—	—	+	+	+	+
6.69				—	+	—	—	—	+
6.63				—	+	—	—	—	+
6.58	+	+	+	—	—	+	+	+	+

^a Sensory descriptors not showing any significant correlation with NMR signal are excluded.

^{*}and ^{*} signs indicate positive and negative correlations ($R^2 > 0.5$), respectively.

^{*}ppm reported in bold are relative to the most intense signals.

^{**}alk stands for alkyl chain.

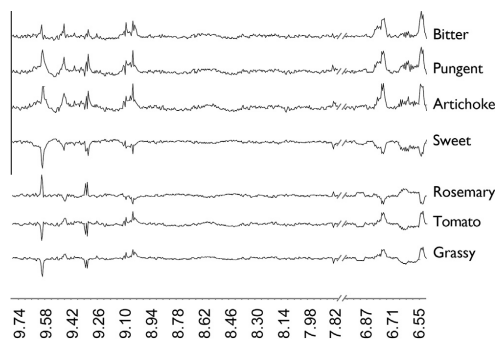


Fig. 6. Loading plot of the predictive component of OPLS models with $Q^2 > 0.4$.

anticorrelation to sweet taste ($R < -0.6$), which displays the inverse profile. Tomato and rosemary tastes also show a strong anticorrelation ($R = -0.75$) and display inverse profiles relative to each other. Grassy and tomato taste display similar profiles without being significantly correlated.

In order to more precisely determine which signals are correlated to which sensory descriptor, the signals displaying correlation above $R^2 > 0.5$ with the OPLS scores for the sensory descriptors were identified (Table 3).

Unfortunately, not all signals reported in Table 3 could be unambiguously assigned and for many of those only a tentative assignment can be done. Thus, signals at δ_H 9.70 could be assigned to the hexenal (Mannina, Patumi, Proietti, Bassi, & Segre, 2001; Sacchi et al., 1997). Signals at δ_H 9.58, δ_H 9.55 and δ_H 6.84 could be assigned to 4-hydroxy-*trans*-alk-2-enal molecules (where *alk* stands for alkyl chain), alk-2,4-dienals and *trans*-alk-2-enals, respectively (Sacchi et al., 1997). Signals around δ_H 9.10 could also be tentatively assigned to protons of the dialdehyd form of secoiridoids, and signals at δ_H 6.5–6.8 to phenyl alcohols moieties (tyrosol and hidroxytyrosol) of oleuropein and ligstroside aglycons (Christophoridou & Dais, 2009). Finally, the signal at δ_H 9.45 is the only one that can be unambiguously assigned to the aldehyde proton of the *trans*-2-hexenal (Mannina et al., 2001; Sacchi et al., 1997).

As expected, this further in-depth inspection of the OPLS data confirmed the qualitative evaluation and allowed us to retrieve much more information, especially regarding the less intense signals. As mentioned above the taste bitter, pungent and artichoke possess a very similar chemical fingerprint. However, they seem to be differentiated by the signals at δ_H 9.21 and 9.31 ppm. Sweet is basically the inverted image of bitter, pungent and artichoke tastes except for the signals at δ_H 9.70, 8.19, 7.88, 6.94, 6.69 and 6.63 ppm. Tomato and grassy are very similar to each other, and they differ only for the signal at δ_H 9.55, 9.11, 9.09, 8.95 and 8.91 ppm. Fruity is almost superimposable to grassy taste except for signals at δ_H 9.55, 9.39, 9.25, 9.21 and 8.95 ppm.

More in general, as for the tentatively assigned signals, the lack of hexenal seems to increase sweet, tomato, grassy and fruity tastes, whereas the increment of its concentration increases the perception of leaf and rosemary tastes. Secoiridoids, *trans*-alk-enals and 4-hydroxy-*trans*-alk-2-enal are in a way related to the sweet/bitter relationship of the EVOO.

Finally, Morales, Alonso, Rios, and Aparicio (1995) proposed that the *trans*-2-hexenal (δ_H 9.45 ppm) correlates to fruity taste of the olive oil. Here we find the same correlation. However, we widen this observation suggesting that this aldehyde is positively correlated to grassy and tomato tastes as well.

4. Conclusions

Extra-virgin olive oil (EVOO) has received increasing attention over the world for their unique nutritional and healthy properties and extraordinary flavor and taste.

Nuclear magnetic resonance spectroscopy (NMR) has been applied to olive oil analysis since 1987 (Sacchi, Addeo, Giudicianni, & Paolillo, 1989). In the last 20 years several applications have been developed mainly applying carbon-13 and proton NMR (Sacchi et al., 1989–1991, 1996–1998; Sacchi, 2001; Mannina et al., 2001, 2003; Zamora, Gomez, Dobarganes, & Hidalgo, 2002; Zamora, Navarro, & Hidalgo, 1994). The recent development of NMR spectrometers (high field, cold-probe) and their performance in the term of both resolution and sensitivity open new perspectives in the application of this powerful analytical technique in the analysis of EVO oil.

The aim of this work was to explore the analytical potentiality of the NMR spectroscopy as “magnetic tongue” in the analysis of extra-virgin olive oil (EVOO), with particular attention to the quantitative measure of minor compounds related to the sensory description. Particularly, the phenol and aldehyde NMR signals allowed a first prediction of sensory characteristics of EVOO.

The reported results are very promising and pave the way to a more careful analysis of other spectral regions on a wider number of oil samples.

Acknowledgements

The authors thank Italo Giudicianni (CIMCF, Naples) for its expertise and kindness and Antonello Paduano for its work. Dr. Raffaele Di Fiore, director of the “Laboratorio Chimico Merceologico” of the Chamber of Commerce of Naples (Italy) is kindly acknowledged. Authors are indebted with Dr. Maria Luisa Ambrosino and to all sensory panelists for their expertise and kindness.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.foodchem.2012.10.135>.

References

- Aishima, T. (2004). Correlating sensory attributes to gas chromatography mass spectrometry profiles and e-nose responses using partial least squares regression analysis. *Journal of Chromatography A*, 1054, 39–46.
- Andrews, P., Busch, J. L. H. C., de Joode, T., Groenewegen, A., & Alexandre, H. (2003). Sensory properties of virgin olive oil polyphenols: Identification of deacetoxy-ligstroside aglycon as a key contributor to pungency. *Journal of Agriculture and Food Chemistry*, 51, 1415–14231.
- Caporale, G., Policastro, S., & Monteleone, E. (2004). Bitterness enhancement induced by cut grass odorant (*cis*-3-hexen-1-ol) in a model olive oil. *Food Quality and Preference*, 15, 219–227.
- Christophoridou, S., & Dais, P. (2009). Detection and quantification of phenol compounds in olive oils by high resolution ^1H nuclear magnetic resonance spectroscopy. *Analytica Chimica Acta*, 633, 283–292.
- Deising, A. K., Stone, D. C., & Thompson, M. (2004). Application of electronic noses and tongues in food analysis. *International Journal of Food Science & Technology*, 6, 587–604.
- Eriksson, L., Johansson, E., Kettaneh-Wold, N., & Wold, S. (1999). Scaling. In *Introduction to multi- and megavariate data analysis using projection methods (PCA & PLS) Umetrics*, 213–225.
- Eriksson, L., Johansson, E., Kettaneh-Wold, N., & Wold, S. (2006). Multi- and megavariate data analysis – Part I: Basic Principles and Applications. *Umetrics*.
- Malmendal, A., Amoresano, C., Trotta, R., Lauri, I., De Tito, S., Novellino, E., et al. (2011). NMR spectrometers as “magnetic tongues”: Prediction of sensory descriptors in canned tomatoes. *Journal of Agriculture and Food Chemistry*, 59, 10831–10838.
- Mannina, L., Dugo, G., Salvo, F., Cicero, L., Ansanelli, G., Calcagni, C., et al. (2003). Study of the cultivar–composition relationship in Sicilian olive oils by GC, NMR, and statistical methods. *Journal of Agriculture and Food Chemistry*, 51, 120–127.

- Mannina, L., Patumi, M., Proietti, N., Bassi, D., & Segre, A. L. (2001). Geographical characterization of Italian extra virgin olive oils using high-field ^1H NMR spectroscopy. *Journal of Agriculture and Food Chemistry*, 49(6), 2687–2696.
- Morales, M. T., Alonso, M. V., Rios, J. J., & Aparicio, R. (1995). Virgin olive oil aroma: Relationship between volatile compounds and sensory attributes by chemometrics. *Journal of Agriculture and Food Chemistry*, 43, 2925–2931.
- Noble, A. C. (2002). Sensory methods of flavor analysis. In A. J. Taylor (Ed.), *Food flavour technology* (pp 252–275). Sheffield, UK: Sheffield Academic Press.
- Roper, S. D. (2006). Cell communication in taste buds. *Cellular and Molecular Life Sciences*, 63, 1494–1500.
- Sacchi, R. (2001). High resolution NMR of virgin olive oil. In *Magnetic resonance in food science—A view to the next century* (pp. 213–226). Cambridge, UK: Royal Society of Chemistry.
- Sacchi, R., Addeo, F., Giudicianni, I., & Paolillo, L. (1989). La spettroscopia di risonanza magnetica nucleare nell'analisi degli oli di oliva. *Rivista Italiana delle Sostanze Grasse*, 56(4), 171–178.
- Sacchi, R., Addeo, F., Giudicianni, I., Paolillo, L. (1990). Applicazione della spettroscopia ^{13}C -NMR alla determinazione di mono-digliceridi ed acidi grassi liberi nell'olio di oliva di pressione e nei rettificati. *Rivista Italiana delle Sostanze Grasse*, 57(5), 245–52.
- Sacchi, R., Paolillo, L., Giudicianni, I., & Addeo, F. (1991). Rapid ^1H -NMR determination of 1,2 and 1,3 diglycerides in virgin olive oils. *Italian Journal of Food Science*, 3(4), 253–262.
- Sacchi, R., Addeo, F., Giudicianni, I., & Paolillo, L. (1992). Analysis of the positional distribution of fatty acids in olive oil triacylglycerols by high resolution ^{13}C -NMR of the carbonyl region. *Italian Journal of Food Science*, 4(2), 117–123.
- Sacchi, R., Patumi, M., Fontanazza, G., Barone, P., Fiordiponti, P., Mannina, L., et al. (1996). A high field ^1H -nuclear magnetic resonance study of the minor components in virgin olive oils. *Journal of the American Oil Chemists Society*, 73(6), 747–758.
- Sacchi, R., Addeo, F., & Paolillo, L. (1997). ^1H and ^{13}C NMR of virgin oil. An overview. *Magnetic Resonance in Chemistry*, 35, 133–145.
- Sacchi, R., Mannina, L., Fiordiponti, P., Barone, P., Paolillo, L., Patumi, M., et al. (1998). Characterization of Italian extra virgin olive oils using ^1H -NMR spectroscopy. *Journal of Agriculture and Food Chemistry*, 46, 3947–3951.
- Savorani, F., Tomasi, G., & Engelsens, S. B. (2010). Icoshift: A versatile tool for the rapid alignment of ^1D NMR spectra. *Journal of Magnetic Resonance*, 202, 190–202.
- Segre, A. L., & Mannina, L. (1997). ^1H NMR study of edible oils. *Recent Research Developments in Oil Chemistry*, 1, 297–308.
- Stone, H., & Sidel, J. L. (1998). Quantitative descriptive analysis: Developments, applications, and the future. *Food Technology*, 52(8), 48–52.
- Taylor, A. J., Rob, S. T., & Linforth, R. S. T. (2003). Direct mass spectrometry of complex volatile and non-volatile flavour mixture. *International Journal of Mass Spectrometry*, 223–224, 179–191.
- Trygg, J., & Wold, S. (2002). Orthogonal projections to latent structures, O-PLS. *Journal of Chemometrics*, 16, 119–128.
- Ward, J. H. Jr., (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 48, 236–244.
- Zamora, R., Gomez, G., Dobarganes, M. C., & Hidalgo, F. J. (2002). Oil fractionation as a preliminary step in the characterization of vegetable oils by high-resolution ^{13}C NMR spectroscopy. *Journal of the American Oil Chemists Society*, 79, 261–266.
- Zamora, R., Navarro, J. L., & Hidalgo, F. J. (1994). Identification and classification of olive oils by high-resolution ^{13}C nuclear magnetic resonance. *Journal of the American Oil Chemists Society*, 71, 361–364.
- Zanor, M. I., Rambla, J., Chaib, J., Steppa, A., Medina, A., Granell, A., et al. (2009). Metabolic characterization of loci affecting sensory attributes in tomato allows an assessment of the influence of the levels of primary metabolites and volatile organic contents. *Journal of Experimental Botany*, 60, 2139–2154.

Supplementary Material

Table S1. Quantitative Descriptive Analysis results.

	fruity	leaf	grassy	bitter	pungent	sweet	almond	artichoke	apple	tomato	rosemary
1	4.5	3.5	1.5	4.5	5.5	3	3	3	1.8	3.3	0
2	5.5	3	3.8	3.5	5.0	3.5	3.5	3.5	2	4	0
3	6.0	4.1	5.1	4.4	4.8	4.3	3.5	3.0	3.5	4.5	0
4	7.0	4.0	5.5	5.3	5.0	3.5	2.8	4.3	3.0	5.5	0
5	6.1	3.3	5	4	4.5	3.9	4.3	3.3	4.1	0	0
6	4	2.5	2.2	3.4	3.9	4.2	3.38	1	2	0	2.5
7	4.8	3.2	3.5	5.0	4.9	3.8	3.5	1.7	3.0	0.0	4.0
8	6	1.7	4.3	4	4.5	4	4.5	2.5	2.3	4.3	0
9	5.8	3.5	4.5	6	5	3	3.1	3.1	3	3.5	0
10	4.7	3.8	3.3	3.4	2.5	3.8	3.5	1.5	3.5	2	0
11	4	1.3	2.8	3	3.5	5.5	3.5	2.4	1.3	3	0
12	5.5	4.5	3.5	5.0	4.5	3.0	3.5	2.3	0	0	4.0
13	7.0	4.5	5.0	6.0	5.2	2.5	2.3	3.5	0	4.2	0
14	6.1	2.8	4.7	5.6	4.8	2.8	4	2.5	2.5	4	0
15	4.5	1.3	3.5	2.5	3	5.3	3.5	0	0	0	3
16	4	1	3.2	3.5	3.6	4.2	3.1	1	2.5	2	0
17	5.5	3.5	4	3.2	3.9	5.0	3.5	0	3.4	2	0
18	6.0	5.0	2.3	3.6	3.5	5.0	2.6	2.0	0	0	5.0

Table S2. Cumulated R^2 and Q^2 values for each variable in the PCA model.

Taste	$R^2(\text{cum})^*$	$Q^2(\text{cum})^{**}$
fruity	0.67	0.41
leaf	0.62	0.25
grassy	0.58	0.22
bitter	0.80	0.68
pungent	0.66	0.45
sweet	0.64	0.40
almond	0.48	-0.07
artichoke	0.71	0.54
apple	0.56	0.12
tomato	0.64	0.37
rosemary	0.75	0.32

* $R^2(\text{cum})$ indicate how well the variation of the variable is explained.

** $Q^2(\text{cum})$ indicate how well the variable can be predicted.

Paper III

Pagano B., **Lauri I.**, De Tito S., Persico G., Chini M. G.,
Malmendal A., Novellino E., Randazzo A.

Use of NMR in profiling of cocaine seizures

Forensic Science International, 2013, 231, 120–124



Use of NMR in profiling of cocaine seizures



Bruno Pagano^a, Ilaria Lauri^a, Stefano De Tito^a, Guido Persico^b, Maria Giovanna Chini^c, Anders Malmendal^d, Ettore Novellino^a, Antonio Randazzo^{a,*}

^a Department of Pharmacy, University of Naples "Federico II", via D. Montesano 49, 80131 Napoli, Italy

^b Laboratorio Indagini Chimiche, Gabinetto Interregionale di Polizia Scientifica per Campania e Molise, via Medina 1, 80133 Napoli, Italy

^c Department of Pharmacy, University of Salerno, via Ponte don Melillo, 84084 Fisciano (SA), Italy

^d Department of Biomedical Sciences, Faculty of Health Sciences, University of Copenhagen, DK-2200 Copenhagen N, Denmark

ARTICLE INFO

Article history:

Received 13 February 2013

Received in revised form 15 April 2013

Accepted 17 April 2013

Available online 17 May 2013

Keywords:

NMR

Multivariate analysis

Cocaine

Profiling

ABSTRACT

Cocaine is the most widely used illicit drug, and its origin is always the focus of intense investigation aimed at identifying the trafficking routes. Since NMR represents a unique methodology for performing chemical identification and quantification, here it is proposed a strategy based on ¹H NMR spectral analysis in conjunction with multivariate analysis to identify the chemical "fingerprint" of cocaine samples, and to link cocaine samples based on this information. The most relevant spectral regions containing the fingerprint have been identified: δ_H 0.86–0.96, 1.50–1.56, 5.90–5.93, 6.48–6.52, 7.31–7.34, 7.61–7.63, 7.68–7.72 ppm. The strategy has been applied on samples seized in different times and places in Naples (Italy). The chemical "fingerprint" depend on what plant they were extracted from, where it was cultivated, and which procedures were used for extraction and purification, thus adding significant information in the process toward identification of the trafficking routes for this drug.

© 2013 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Cocaine is the purification product of coca paste, an extract of the leaves of the coca bush, where it is found as a natural alkaloid. In addition to cocaine, which is the main alkaloid, coca paste contains also small percentages of other compounds, mostly alkaloids [1], here named "minor components". The amount and the kind of these compounds depend on the specific cultivar and on the environment where the plant has been cultivated. Interestingly, the procedures used for extraction and purification of the cocaine are not totally efficient, so that small amount of minor components are found also in trafficked cocaine samples. Therefore, the presence of these minor compounds is very dependent of the "history" of the cocaine sample and they can be considered a "fingerprint" of that sample. A careful analysis of this fingerprint can provide very important information about cocaine origin, or more simply allow the identification of a given consignment as well as the area and period in which that consignment has been trafficked. Minor components may have forensic significance, since the origin of cocaine samples can, in principle, be determined through the analytical determination of their presence or absence [2].

Although a large array of techniques for the analysis and identification of cocaine is available, the GC–MS technique currently dominates forensic analysis [3]. Nevertheless, sometimes this technique requires a derivatization and a chromatographic separation. Unfortunately, not all molecules are amenable to derivatization, and once derivatized, analytes could not be sufficiently volatile or stable for GC separations. Nuclear magnetic resonance (NMR) spectroscopy represents one of forensic sciences' most versatile tools [4], since it represents a unique methodology for performing identification and quantification at the same time without any derivatization. In addition, NMR does not require a high purity reference standard for accurate quantitation of the target compound and it does not have a medium which can lead to solute adsorption effects and imprecision of analysis.

The determination of the origin of illicit cocaine samples seized by law enforcement is always the focus of intense investigation aimed at identifying the trafficking routes. In fact, drug trafficking is an acute problem compounded by judicial rules and by difficulties in the exchange of data, due to data protection regarding people and investigative processes. However, seized drugs can be easily compared in order to provide information without interfering with the normal investigative processes. The considerable quantities of seized drugs allow data collection that can form a useful basis for such procedures. The results of such comparisons may provide key evidence in the investigations.

In the present study, a strategy based on ¹H NMR spectral analysis applied in conjunction with multivariate analysis is

* Corresponding author. Tel.: +39 081 678514; fax: +39 081 678514.

E-mail addresses: antonio.randazzo@unina.it, antranda@unina.it (A. Randazzo).

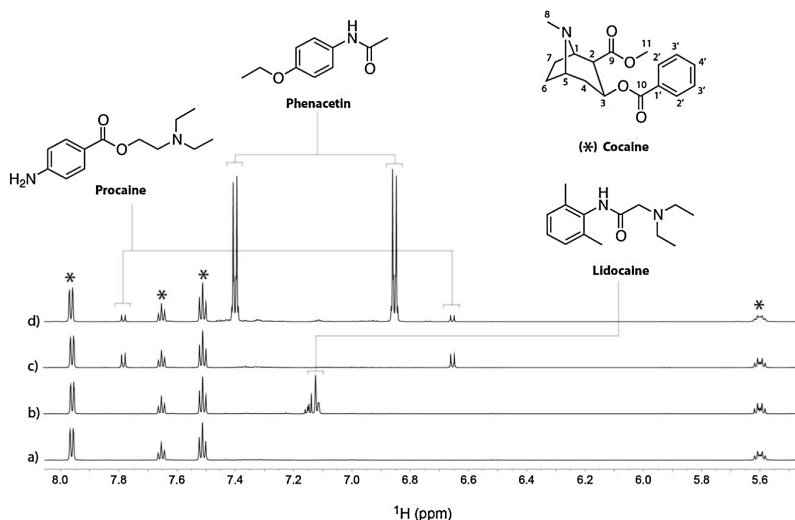


Fig. 1. Comparison of an expanded region of four samples of seized cocaine. (a) Example of a sample of pure cocaine. Asterisks indicate cocaine signals. (b) Sample having lidocaine as cutting agent. (c) Sample having procaine as cutting agent. (d) Sample having procaine and phenacetin as cutting agents.

proposed. This investigation allows determining linkage between cocaine samples seized at different times and different places in Naples (Italy) to indicate common origin or distribution channel, which, combined with traditional information, can improve investigation strategies.

2. Materials and methods

2.1. Materials

54 chlorohydrate cocaine samples were seized by the Police Department of Naples in different areas of Naples county during the year 2006.

2.2. NMR spectroscopy

2 mg of each sample were dissolved in 600 μ L CD₃OD and placed in 5 mm NMR tubes without any preliminary treatment. ¹H NMR spectra were acquired at 25 °C with a 700 MHz Varian Unity Inova spectrometer using a 5 mm ¹H/¹³C/¹⁵N triple resonance probe. The ¹H NMR measurements were carried out with 1000 transients and 32k complex data point. In order to retrieve quantitative information, the recycle time was set to 5 s, and a 45° pulse angle was used.

2.3. GC–MS

Seized samples were dissolved in absolute ethyl alcohol (Carlo Erba Reagenti, Italy) to a concentration of 1 mg/mL. GC was performed with an Agilent 6890A gas

chromatograph equipped with a quadrupole mass-selective detector (MSD) Agilent 5973N (Agilent Technologies). The MSD was operated in the electron ionization (EI) mode, with an ionization potential of 70 eV, a scan range of 40–450 amu. The GC was fitted with a 30 m \times 0.25 mm ID fused-silica capillary column coated with 0.25 μ m 5% diphenyl–95% dimethylpolysiloxane stationary phase (HP5–ms, J&W). The oven temperature was programmed as follows: initial temperature, 180 °C; initial hold, 2 min; program rate, 10 °C/min, 270 °C; hold, 2 min; 10 °C/min; 300 °C final temperature; hold, 2 min (run time 18 min). Column flow: 1 mL/min. The injector was operated in the split mode (10:1) at 280 °C. Injection volume 1 μ L. The MSD source was operated at 230 °C.

2.4. NMR data reduction and processing

The spectra were processed using iNMR software (www.inmr.net). An exponential line broadening of 0.5 Hz was applied to the free-induction decay prior to Fourier transformation. All spectra were referenced relative to external sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS), phased and baseline corrected. Data reduction was accomplished by dividing the spectrum into 0.005 ppm regions (bins) over which the signal was integrated to obtain the signal intensity. This was done in order to minimize artifacts due to small variations in peak position and to reduce the number of variables without losing spectral information. Only the spectral region between 0.0 and 8.5 ppm (excluding the region around the solvent signals between 3.29–3.33 and 4.79–4.83 ppm) was considered for the study (see Section 3). The integrals were normalized to the integral of the cocaine triplet at δ_H 7.65 ppm (H4') in order to suppress trivial separation based on variations in the amount of sample.

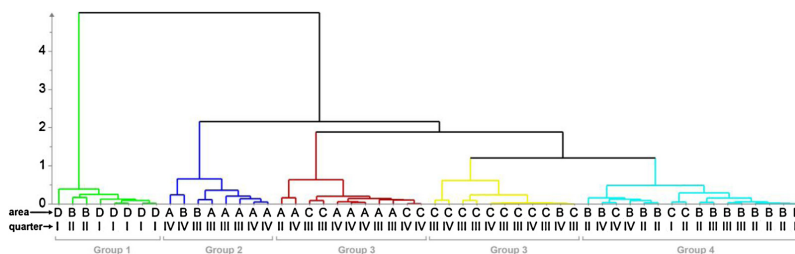


Fig. 2. HCA dendrogram based on NMR data, showing similarities between samples. Samples falling within the same group are indicated with the same color. The dendrogram is calculated with Ward clustering method and sorted by size. Area (A, B, C and D) and quarter of the year (I, II, III and IV) of seizure are also reported. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Spectral ranges containing clustering information.

Range	ppm
1	0.86–0.96
2	1.50–1.56
3	5.90–5.93
4	6.48–6.52
5	7.31–7.34
6	7.61–7.63
7	7.68–7.72

2.5. Statistical analysis

Statistical analysis was performed using Simca-P 13.0 (Umetrics, Umeå, Sweden). The number of principal components (PCs) in the Principal Component Analysis (PCA) [5] was determined by cross-validation. The model used was estimated using all 54 samples. The dataset was pareto-scaled [6], i.e. each of the values in each bin were divided by the square root of the standard deviation for all the values in the bin, and centered. Hierarchical cluster analysis (HCA) was carried out by using the Ward clustering method [7]. Partial least-squares regression discriminant analysis (PLS-DA) was performed to explore the ability of the NMR data to discriminate between classes of samples [8]. The quality of the models was described by R^2 and Q^2 values. R^2 is defined as the proportion of variance in the data explained by the models and indicates the goodness of fit. Q^2 is defined as the proportion of variance in the data predictable by the model and indicates predictability [8].

The model validation was performed using the permutation test [9], in which a total of 400 models were calculated by randomizing the order of Y variables in the corresponding PLS-DA models. The obtained Q^2 and R^2 values, describing the predictive ability and the reliability of the fitting, respectively, were plotted and compared with the Q^2 and R^2 values obtained from the real model.

Statistical Total Correlation Spectroscopy (STOCSY) analysis was performed on the whole ^1H NMR data set of the 54 samples, to obtain the correlations among the constituents, using a cutoff value $r > 0.9$ for correlation coefficient.

3. Results and discussion

Each seized sample has been analyzed by ^1H NMR without any preliminary treatment. It was easy to detect the main signals belonging to the cocaine hydrogens (see Table S1 in Supplemental Materials). Nevertheless, other signals were present in the spectra. These signals could be attributed to the minor components (i.e. norcocaine, tropacocaine, *cis*- and *trans*-cinnamoylcocaine, ecgonine and ecgonidine methyl esters) coming from the extraction and purification processes of the cocaine and/or cutting agents (adulterant/diluent) that could be mixed to the cocaine samples before drug dealing (Fig. 1). The analysis of cutting agents can be used as an investigative support, since they can be easily detected and quantified by a simple ^1H NMR spectrum. For example, their analysis can be used to identify the specific cutting agents being used in a given area (or period) by clandestine laboratories. However, since the main goal of this investigation is to propose a strategy to determine a common origin and/or a distribution channel of the illicit drug, adulterants or diluents are not useful for our purpose since different local pushers may have cut the same cocaine consignment in different ways. On the other hand, as mentioned above, the entire set of signals belonging to the minor components can be considered a “fingerprint” of that sample. Thus, a careful analysis of this fingerprint can provide very important information about cocaine origin, or more simply allow the identification of the areas and period in which a given consignment of drug has been trafficked. Therefore, the first step of the work was the identification of spectral regions of the ^1H NMR spectrum containing information related only to the minor components, without interference from cocaine and adulterant/cutting agent signals.

With this aim, the STOCSY analysis method has been applied [10]. This technique can be used for determining metabolic connectivity between different molecules as well as for structural

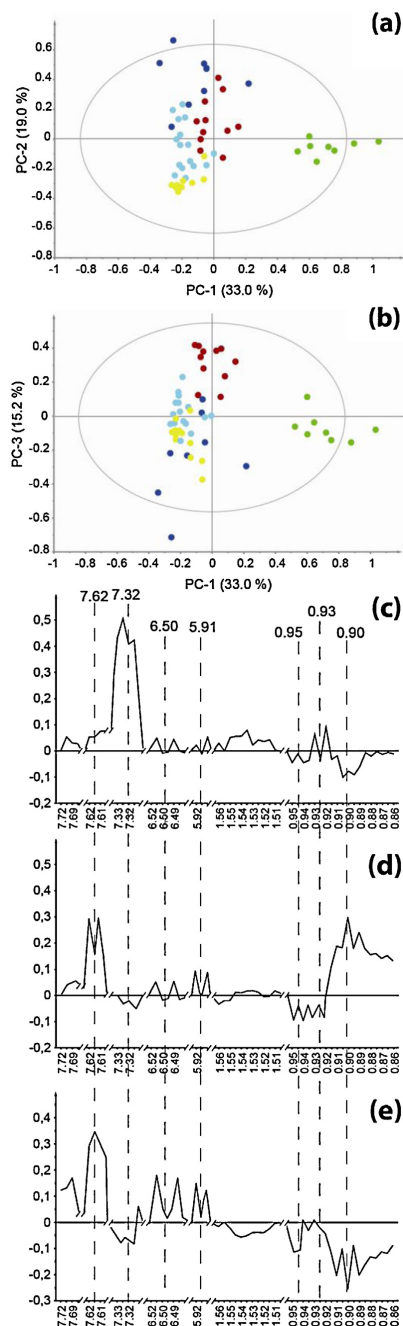


Fig. 3. PCA score (panels a and b) and loading (panels c–e) plots. Panels (a) and (b) show the PC1–PC2 and PC1–PC3 score plots. Panels (c), (d) and (e) show the PC1, PC2 and PC3 loading plots respectively. Samples are colored according to HCA analysis in Fig. 2.

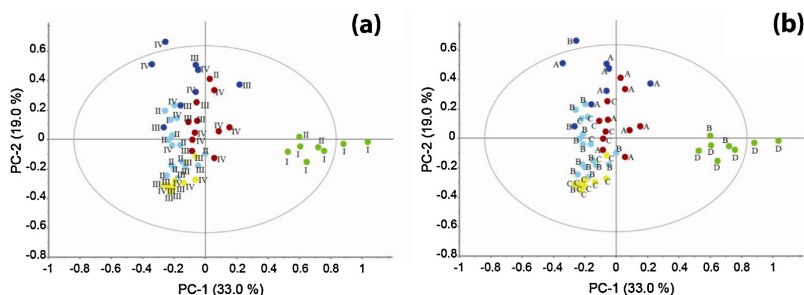


Fig. 4. PCA score-plots annotated with (panel a) year quarter and (panel b) area of seizure.

assignment in NMR spectra of complex mixtures. STOCY takes advantage of the multicollinearity of the intensity variables in a set of spectra (in this case the 54 ^1H NMR spectra) to generate a pseudo-2D NMR spectrum, which displays correlation among the intensities of various peaks across the whole spectrum. In the STOCY plot (see Fig. S1 in Supplemental Materials), correlations occurring from a single molecule (structural correlations) are present, as well as the intermolecular ones (non structural correlations) leading to the identification of substances whose signal intensities increase and decrease simultaneously with the first molecule. Here, only the structural correlations were considered. Thus, it was possible to identify all the signals of cocaine and cutting agents with the help of literature data [11]. The signals for the following agents were detected: lidocaine, phenacetin, diltiazem, sugars, procaine, MDMA (3,4-methylenedioxy-N-methylamphetamine), paracetamol and caffeine. Hence, all the spectral regions in which those signals were present were excluded, and all the regions that only contain signals of the minor (fingerprint) components were considered in the analysis. Spectral regions without signals have also been excluded. Overall, only seven spectral intervals (Table 1) containing signals of minor components were identified.

In order to improve the comparability of the spectra, they have been normalized. Since the amount of the minor components is somehow related to the amount of cocaine (which is the main component), the 7 spectral intervals of each sample were normalized with respect to the area of a reference cocaine peak. Particularly, the cocaine triplet at δ_{H} 7.65 ppm (H_4') has been used, since it does not overlap with other signals of any of the 54 samples. Such normalization provided the relative quantity of the fingerprint constituents for each sample. The resulting data matrix, with columns representing the normalized signal intensities and rows representing the different samples, was then used for the multivariate analysis.

One of the goals of this investigation is to find similarities among samples, in order to cluster them and to determine the number of drug consignments trafficked during a given period of sampling. Thus, hierarchical cluster analysis (HCA) has been performed on the data matrix. This analysis finds clusters of samples based on intensities of the signals present in the selected spectral regions. Before clustering, all the dataset have been pareto-scaled and centered. Pareto scaling, i.e. dividing the (mean-centered) variables (points) by the square root of their standard deviation, is applied to reduce the weight of the high-intensity components, while reducing the weight of noise in the model. The resulting HCA dendrogram (Fig. 2) shows that the 54 samples are gathered in 5 main groups that can be considered as 5 different drug consignments. In order to get insights about the reasons for this clustering, principal component analysis (PCA) was performed

on the same dataset (Fig. 3). Three principal components (PCs), accounting for 67% (PC1 33%, PC2 19%, PC3 15%) of the variation were identified. These numbers are in the usual range for NMR data. The explained variance is not so high because the NMR data contains noise. The plots of their scores (Fig. 3a and b) show the positioning of the samples according to their chemical characteristics and allowed the identification of the most important NMR signals for sample differentiation. PCA indicates that group 1 substantially differs from groups 2, 3, 4 and 5 along PC1. According to the PC1 loading plot (Fig. 3c), group 1 differs from the other groups mainly for the presence of tropacocaine, whose hydrogens resonate at δ_{H} 7.32 ppm [12]. The presence of this minor component in the samples belonging to group 1 has been also established by GC–MS analysis performed on the same samples. On the other hand, groups 2, 3, 4 and 5 are separated along PC2. The PC2 loading plot indicates that these groups mainly differ for the content of *cis*- (δ_{H} 7.62 and 5.91 ppm) and *trans*-cinnamoylcocaine (δ_{H} 6.50 ppm) [13]. Furthermore, the four groups are also separated by a different content of unassigned molecules having signals at δ_{H} 0.95, 0.93 and 0.90 ppm. The samples with high PC2 scores are characterized by higher concentrations of *cis*- and *trans*-cinnamoylcocaine, and of the molecule having signal at δ_{H} 0.90 ppm, and lower concentrations of the molecules having signals at δ_{H} 0.95 and 0.93 ppm. PC3 is also governed by the same resonances. Indeed, the loading plots of PC2 and PC3 (Fig. 3d and e) are very similar except that the signal at 0.90 ppm has the same sign as the signals at 0.93 and 0.95 ppm in PC3.

In order to determine how well the identified consignments of drug can be predicted by NMR, predictive models for the different groups were computed, using the Projection to Latent Structures Discriminant Analysis (PLS-DA) (Fig. S2 (a and b) in Supplemental Materials). The model gave prediction scores ($Q^2(\text{cum}) > 0.5$) for all five groups (Fig. S2 (c) in Supplemental Materials). The validity of the PLS-DA model was assessed with the use of the permutation test, in which a total of 400 models were calculated using randomly permuted Y variables (Fig. S2 (d) in Supplemental Materials). The obtained Q^2 and R^2 values, describing the predictive ability and the reliability of the fitting, respectively, were plotted and compared with the Q^2 and R^2 values obtained from the real model. The substantial decrease of both parameters Q^2 and R^2 (vertical axis interception point of the Q^2 and R^2 regression line resulted both with negative values) enforced the statistical validity of the obtained PLS-DA model. In order to further validate the PCA and PLS-DA models, five PCAs and five PLS-DAs additional models have been computed. Each of these models has been computed holding out 10 randomly selected samples at the time. All models were characterized by high predictive coefficients ($Q^2 > 0.5$), indicating that the original models well describe the distribution of the samples.

In order to retrieve investigative information, all the samples have been labeled according to place (Figs. 2 and 4b) and date (Figs. 2 and 4a) of seizure. Based on information from Naples Police Department, four macroareas (A, B, C and D) of Naples county have been considered, in a way that each macroarea is controlled by a single criminal organization (clan). The samples have also been labeled according to the quarter of the year in which they were seized (I, II, III and IV) (Fig. 4a). These labels reveal a very interesting structure of the data. For example, samples of group 1 (green) have been seized mainly in area D of the county. Analogously, samples of groups 2 (blue), 3 (red), 4 (yellow) and 5 (cyan) are seized mainly in areas A, A/C, C, and B, respectively. This may suggest that each clan stocks up with drug consignments independently from the others and that it actually controls a given area in an almost exclusive manner. It is also possible to monitor the samples that, for some reason, move from the area where most of the consignment to which they belong has been sold. For example, this could be due to a reselling of the cocaine sample or to a simple moving of a consumer from the area of purchase to the one of consumption.

The consignments were also trafficked in different periods. Thus, group 1 (green) have been sold mainly in the first (I) quarter, with only a couple of samples in the second one (II) (Fig. 4a). Groups 2, 3 and 4 have been trafficked in the third (III) and forth (IV) quarters and, finally, group 5 (blue) has been sold all year long.

4. Conclusion

In the course of a police investigation on a cocaine distribution network, the investigators obtain large amounts of circumstantial information. On the basis of those data, they infer and define links that may exist between different persons active within a distribution network. Linkages revealed using such traditional methods of investigation can be corroborated and even substantiated by the detection of chemical links, which can, in turn, be used to reveal previously undetected investigative links. In this frame, the potentiality of the NMR in combination to the statistical multivariate analysis is demonstrated to potentially ameliorate for police investigations. In particular, the spectral regions that, at least in this investigation, can be used for grouping of seized cocaine samples have been defined: δ_H 0.86–0.96, 1.50–1.56, 5.90–5.93, 6.48–6.52, 7.31–7.34, 7.61–7.63, 7.68–7.72 ppm. These regions were very useful in providing information about place and date of trafficking. The statistical multivariate analysis allowed also the identification of key minor components responsible of this grouping.

The results reported here indicate that the proposed analysis can assist tactically (evidential/judicial) and strategically (intelligence) the investigators. This can contribute to the establishment of distribution and/or trafficking links between multiple seized

samples that have been obtained at different locations or in the possession of different individuals. Furthermore, the fact that the relative ratios of the minor components in coca leaf are closely associated with plant varietal, cultivar and agronomic differences can be exploited for the assignment of geographical origin, at least when suitable authentic databases is available.

Acknowledgements

Authors are grateful to the Police Department of Naples for helping us with its experience. Authors are also indebted with Dr. Loredana Randazzo for her very important contribution in developing Matlab scripts.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.forsciint.2013.04.028>.

References

- [1] D.S. Isenschmid, Cocaine, in: Wiley Encyclopedia of Forensic Science, John Wiley and Sons, New York, 2009.
- [2] J.M. Moore, J.F. Casale, Cocaine profiling methodology – recent advances, *Forensic Sci. Rev.* 10 (1998) 13–46.
- [3] J.M. Moore, J.F. Casale, In-depth chromatographic analyses of illicit cocaine and its precursor, coca leaves, *J. Chromatogr. A* 674 (1994) 165–205.
- [4] P.A. Hays, Proton nuclear magnetic resonance spectroscopy (NMR) methods for determining the purity of reference drug standards and illicit forensic drug seizures, *J. Forensic Sci.* 50 (2005) 1342–1360.
- [5] L. Eriksson, E. Johansson, N. Kettaneh-Wold, S. Wold, Multi- and Megavariate Data Analysis – Part I: Basic Principles and Applications, second ed., Umetrics, Sweden, 2006.
- [6] L. Eriksson, E. Johansson, N. Kettaneh-Wold, S. Wold, Scaling, in: Introduction to Multi- and Megavariate Data Analysis using Projection Methods (PCA & PLS), Umetrics, Sweden, 1999, pp. 213–225.
- [7] J.H. Ward Jr., Hierarchical grouping to optimize an objective function, *J. Am. Stat. Assoc.* 48 (1963) 236–244.
- [8] E. Szymanska, E. Saccenti, A.K. Smilde, J.A. Westerhuis, Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies, *Metabolomics* 8 (2012) 3–16.
- [9] J.A. Westerhuis, H.C.J. Hoefsloot, S. Smit, D.J. Vis, A.K. Smilde, E.J.J. van Velzen, J.P.M. van Duinhoven, F.A. van Dorsten, Assessment of PLS-DA cross validation, *Metabolomics* 4 (2008) 81–89.
- [10] O. Cloarec, M.E. Dumas, A. Craig, R.H. Barton, J. Trygg, J. Hudson, C. Blancher, D. Gauguier, J.C. Lindon, E. Holmes, J. Nicholson, Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR Data Sets, *Anal. Chem.* 77 (2005) 1282–1289.
- [11] C.J. Groombridge, NMR spectroscopy in forensic science, *Annu. Rep. NMR Spectrosc.* 32 (1996) 215–297.
- [12] S.P. Singh, D. Kaufman, V.I. Stenberg, Nitrogen photochemistry. Cocaine and its model compounds, *J. Heterocycl. Chem.* 16 (1979) 625–631.
- [13] (a) J.M. Moore, Identification of cis- and trans-cinnamoylcocaine in illicit cocaine seizures, *J. AOAC* 56 (1973) 1199–1205;
(b) J.F. Casale, E.L. Pinero, E.M. Corbeil, Isolation of cis-cinnamoylcocaine from crude illicit cocaine via alumina column chromatography, *Microgram J.* 4 (2006) 37–41.

Supplemental Materials

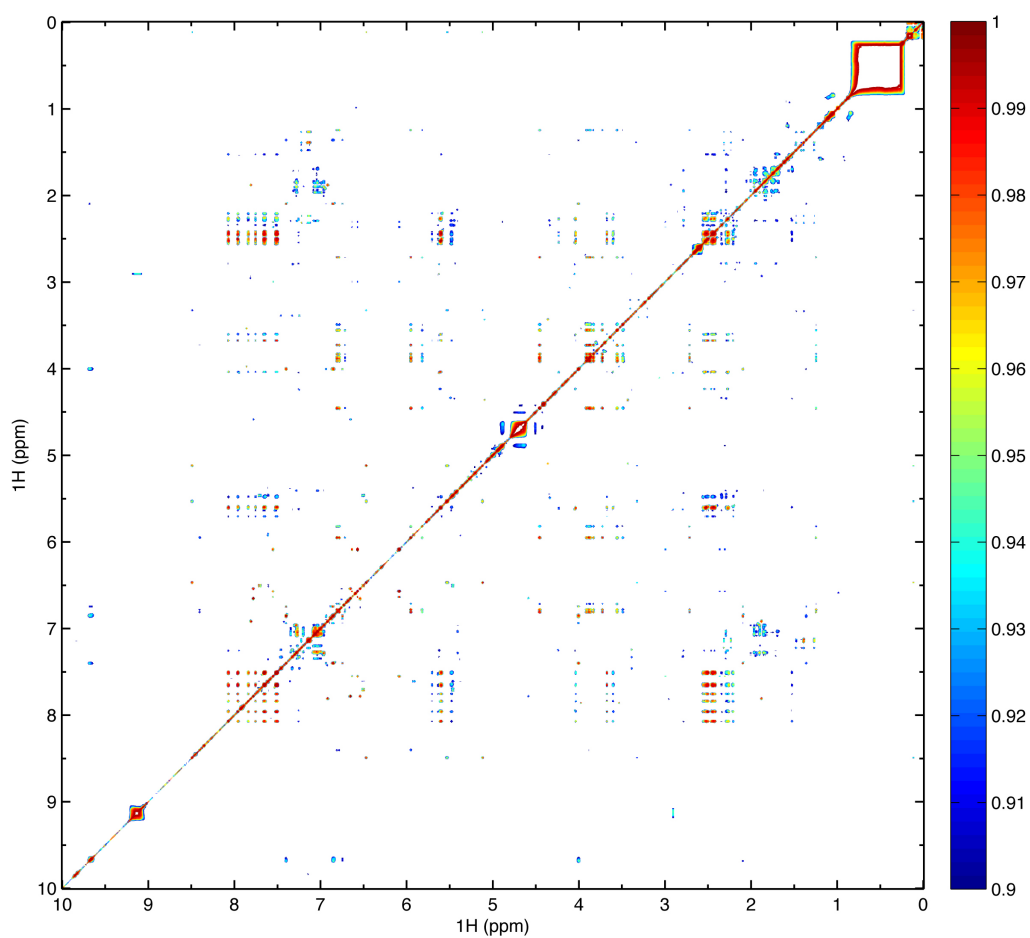


Fig. S1. STOCSY plot for the ^1H NMR spectra of cocaine samples. Correlation matrix is calculated from 54 spectra and is plotted as a contour plot with $r > 0.9$.

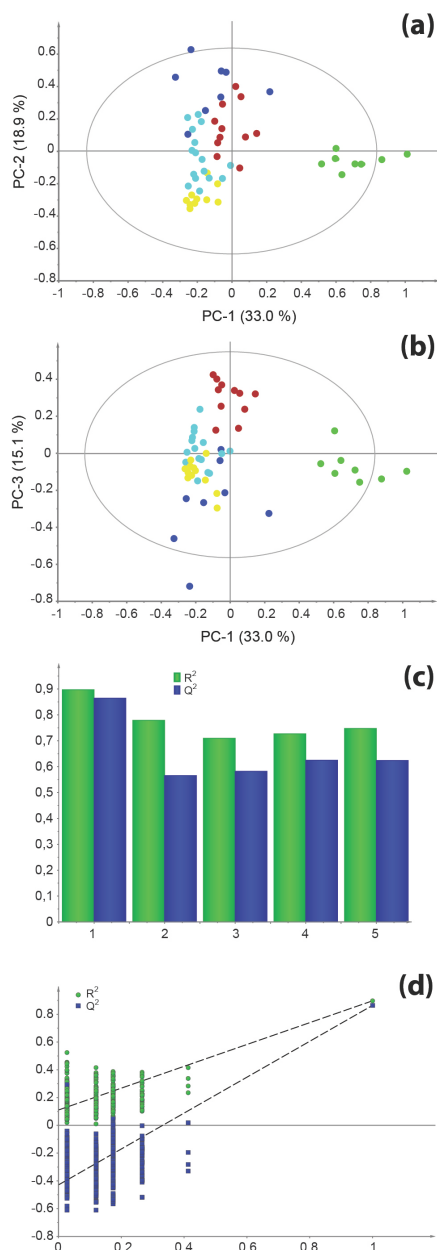
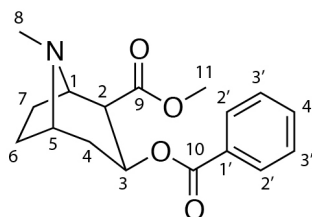


Fig. S2. (a, b) PC1-PC2 and PC1-PC3 score plots of PLS-DA. (c) Cumulated R^2 and Q^2 values for each variable of the training set (Y for PLS). R^2 indicates how well the variation of a variable is explained, while Q^2 indicates how well a variable can be predicted. The latter is estimated by cross validation. (d) Permutation test in which 400 models were calculated by randomizing the order of Y variables. The obtained Q^2 and R^2 values, describing the predictive ability and the reliability of the fitting, respectively, were plotted and compared with the Q^2 and R^2 values obtained from the real model.

Table S1. ^1H and ^{13}C NMR assignments of cocaine.



position	^1H (ppm)	^{13}C (ppm)
1	4.27	65.2
2	3.60	47.2
3	5.59	64.8
4	2.43	33.7
5	4.07	64.4
6	2.25, 2.46	23.5
7	2.23, 2.53	24.6
8	2.92	39.4
9	–	173.6
10	–	166.2
11	3.66	53.3
1'	–	130.0
2'	7.95	130.1
3'	7.50	129.5
4'	7.65	134.6

SUPPLEMENTAL MATERIAL

Trotta R., De Tito S., **Lauri I.**, La Pietra V.,
Marinelli L., Cosconati S., Martino L.,
Conte M.R., Mayol L., Novellino E.,
Randazzo A.

A more detailed picture of the interactions between virtual
screening-derived hits and the DNA G-quadruplex: NMR,
molecular modelling and ITC studies

Biochimie, 2011, 93, 1280-1287

Cosconati S., Rizzo A., Trotta R., Pagano B., Iachettini S., De Tito S.,
Lauri I., Fotticchia I., Giustiniano M., Marinelli L., Giancola C.,
Novellino E., Biroccio A., A. Randazzo

Shooting for Selective Druglike G-Quadruplex Binders:
Evidence for Telomeric DNA Damage and Tumor Cell Death

Journal of Medicinal Chemistry, 2012, 55 (22), 9785-9792.



Contents lists available at ScienceDirect

Biochimie

journal homepage: www.elsevier.com/locate/biochi



Research paper

A more detailed picture of the interactions between virtual screening-derived hits and the DNA G-quadruplex: NMR, molecular modelling and ITC studies

Roberta Trotta^a, Stefano De Tito^b, Ilaria Lauri^b, Valeria La Pietra^b, Luciana Marinelli^b, Sandro Cosconati^b, Luigi Martino^c, Maria R. Conte^c, Luciano Mayol^a, Ettore Novellino^b, Antonio Randazzo^{a,*}

^a Dipartimento di Chimica delle Sostanze Naturali, Università degli Studi di Napoli "Federico II", via D. Montesano 49, I-80131 Napoli, Italy

^b Dipartimento di Chimica Farmaceutica e Tossicologica, Università degli Studi di Napoli "Federico II", via D. Montesano 49, I-80131 Napoli, Italy

^c Randall Division of Cell and Molecular Biophysics, King's College London, New Hunt's House, Guy's Campus, London SE1 1UL, UK

ARTICLE INFO

Article history:

Received 27 March 2011

Accepted 23 May 2011

Available online 30 May 2011

Keywords:

Virtual screening

NMR titration

Brominated G-quadruplex

Isothermal titration calorimetry

ABSTRACT

The growing amount of literature about G-quadruplex DNA clearly demonstrates that such a structure is no longer viewed as just a biophysical strangeness but it is instead being considered as an important target for the treatment of various human disorders such as cancers or venous thrombosis. In this scenario, with the aim of finding brand new molecular scaffolds able to interact with the groove of the DNA quadruplex [d(TGGGGT)]₄, we recently performed a successful structure-based virtual screening (VS) campaign. As a result, six molecules were found to be somehow groove binders. Herein, we report the results of novel NMR titration experiments of these VS-derived ligands with modified quadruplexes, namely [d(TGG^{Br}GGT)]₄ and [d(TGGGG^{Br}T)]₄. The novel NMR spectroscopy experiments combined with molecular modelling studies, allow for a more detailed picture of the interaction between each binder and the quadruplex DNA. Noteworthy, isothermal titration calorimetry (ITC) measurements on the above-mentioned compounds revealed that **2**, **4**, and **6** besides their relatively small dimensions bind the DNA quadruplex [d(TGGGGT)]₄ with higher affinity than distamycin A, to the best of our knowledge, the most potent groove binder identified thus far.

© 2011 Elsevier Masson SAS. All rights reserved.

1. Introduction

Telomeres are special structures that adorn the end of all eukaryotic chromosomes. They are an ensemble of proteins (shelterin proteins) and noncoding DNA sequence which consists, in all vertebrates, of 5'-(TTAGGG)_n-3' repeats, followed by a G-rich single-stranded 3'-overhang (G-tail) [1]. Telomeres are fundamental for the protection of chromosomal ends from unwanted recombination and degradation and at the same time they allow the chromosomes ends to be distinguished from chromosome breaks thus avoiding 'repair' processes that would result in chromosome end-to-end fusions. It is well-known that telomeric DNA in normal somatic cells progressively shortens at each round of cell division as a consequence of the inability of DNA polymerase to fully replicate the 3' ends. Thus, when the Hayflick limit is reached, cells stop replicating and enter a senescence phase which precedes apoptosis.

Therefore, telomeric DNA functions also as a biological clock, at least in healthy cells. Conversely, cancer cells are able to maintain telomeric DNA length constant through the expression of telomerase enzyme which catalyses the synthesis of G-rich repeats at 3' [2]. Telomerase is indeed over-expressed in the majority of human cancers and now it is a validated target for the search of novel effective antineoplastic agents [3]. Besides, the well-known telomerase inhibitors which are specific for the catalytic subunit, a novel class of telomerase inhibitors, known as G-quadruplex stabilisers, is emerging [4], and they arise from the evidence that the G-tail of the telomeric DNA can form unusual structures termed G-quadruplexes. G-quadruplex structures comprise stacks of G-tetrads, which are the planar association of four guanines in a cyclic Hoogsteen hydrogen-bonding arrangement [5]. The formation of quadruplex structures at telomeric DNA level results in telomerase inhibition since quadruplex DNA is not recognised by the single-stranded RNA component of the telomerase enzyme. But, more importantly, the quadruplex structure can be recognised itself as DNA damage signal thus instantly invoking apoptosis [6]. At present, a growing number of quadruplex binders exist and they

* Corresponding author. Tel.: +39 081 678514; fax: +39 081 678552.
E-mail address: antonio.randazzo@unina.it (A. Randazzo).

mostly stack on the terminal quartet surface (end-stackers) [7]. The first evidence that the grooves of a quadruplex structure can be recognised by an organic molecule came approximately three years ago from our own NMR work revealing that, distamycin A can bind in a dimeric form to the two opposite grooves of the quadruplex [8]. Whereas the end-stacker ligands are the major part of all known quadruplex binders and their number is growing each day, quadruplex groove binders represent a quite unexplored and valuable field. In fact, grooves in duplex and quadruplex DNA are chemically and conformationally different, and since groove dimensions differ according to the type of quadruplex, groove binders can in principle be selective for a particular quadruplex topology. Thus, with the aim of finding brand new molecular scaffolds able to interact with the groove, starting from the quadruplex structure found in complex with distamycin A, we recently performed an extensive structure-based virtual screening (VS) campaign [9]. As a result, six molecules (**1–6**, Fig. 1) were found to be somehow groove binding agents. Herein, we have performed additional NMR titrations of the six compounds with modified quadruplexes, namely $[d(TGG^{Br}GGT)]_4$ and $[d(TGGGG^{Br}T)]_4$. These NMR spectroscopy experiments combined with molecular modelling studies, provided a more detailed picture of the interactions between each binder and the quadruplex DNA. Noteworthy, isothermal titration calorimetry (ITC) measurements on the above-mentioned compounds revealed that **2**, **4**, and **6** despite their relatively small dimensions, are more effective in groove binding with respect to distamycin A, to the best of our knowledge, the most potent groove binder identified thus far.

2. Material and methods

2.1. Oligonucleotide synthesis

The oligonucleotide d(TGGGGT) was synthesised on a Millipore Cyclone Plus DNA synthesizer using solid phase β -cyanoethyl phosphoramidite chemistry at 15 μ mol scale. Commercially available 5'-DMT-aminoprotected-8-bromodeoxyguanosine-3'-phosphoramidite was used for the preparation of the modified oligonucleotides. The oligomers were detached from the support and deprotected by treatment with concentrated aqueous ammonia at 55 °C for 12 h. The combined filtrates and washings were concentrated under reduced pressure, redissolved in H₂O, analysed and purified by high-performance liquid chromatography (HPLC) on a Nucleogel SAX column (Macherey–Nagel, 1000-8/46); using buffer A: 20 mM KH₂PO₄/K₂HPO₄ aqueous solution (pH 7.0), containing 20% (v/v) CH₃CN; buffer B: 1 M KCl, 20 mM KH₂PO₄/K₂HPO₄ aqueous solution (pH 7.0), containing 20% (v/v) CH₃CN; a linear gradient from 0 to 100% B for 30 min and flow rate 1 ml/min were used. The fractions of the oligomer were collected and successively desalted by Sep-pak cartridges (C-18). The isolated oligomers proved to be >98% pure by NMR.

2.2. Nuclear magnetic resonance experiments

The quadruplex NMR samples were prepared at a concentration of 2 mM (8 mM single strand concentration), in 0.6 ml (H₂O/D₂O

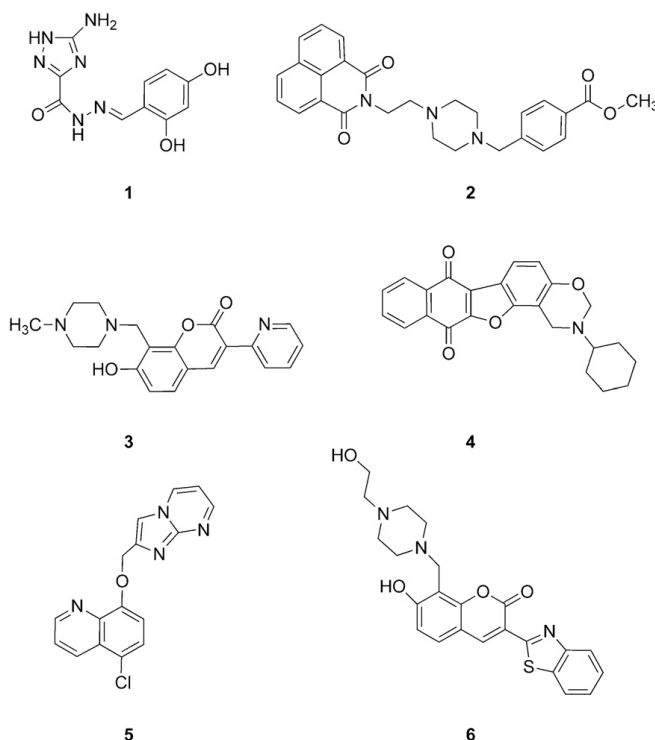


Fig. 1. Structure of the newly identified quadruplex binders as resulted from the virtual screening experiment [9].

9:1) buffer solution having 10 mM KH_2PO_4 , 70 mM KCl, 0.2 mM EDTA, pH 7.0. NMR spectra were recorded with Varian UnityINOVA 700 MHz spectrometer. ^1H chemical shifts were referenced relative to external sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS). 1D proton spectra of the sample in H_2O were recorded using pulsed-field gradient DPFGSE [10,11] for H_2O suppression. Phase-sensitive NOESY spectra [12] were recorded with mixing times of 100 and 200 ms ($T = 25^\circ\text{C}$). Pulsed-field gradient DPFGSE [10,11] sequence was used for NOESY experiments in H_2O .

NOESY experiments were recorded using STATES-TPPI [13] procedure for quadrature detection. The time domain data consisted of 2048 complex points in t_2 and 400 fids in t_1 dimension. A relaxation delay of 1.2 s was used. The NMR data were processed on iMAC running iNMR software (www.inmr.net).

2.3. Molecular modelling

The binding modes of compounds **1–6** were studied by means of docking experiments with the aid of Autodock4 (AD4) [14] and using the $[\text{d}(\text{TGGGGT})_4]$ G-quadruplex DNA structure as deposited in the Protein Data Bank (PDB code 1S45) as macromolecules. The 3D structures of all the compounds were generated with the Maestro Build Panel [15]. For the purpose of docking each molecule has been constructed in the protonation state suggested by the MarvinSketch 5.2.5.1 package (<http://www.chemaxon.com>) using a pH 7.0 accordingly with the NMR titrations. The target DNA structures were prepared through the Protein Preparation Wizard of the graphical user interface Maestro 9.0.211 [15] and the OPLS-2001 force field. Water molecules were removed, hydrogen atoms were added and minimisation was performed until the RMSD of all heavy atoms was within 0.3 Å of the crystallographically determined positions. Then the constructed compounds and DNA structures were converted to AD4 format files using ADT generating automatically all other atom values. In order to allow the ligands to explore all the possible search space, the docking area has been centred on the mass centre of the quadruplex structure and defined by a box large enough to comprise the entire macromolecule. Accordingly, grids points of $84 \times 84 \times 84$ with 0.375 Å spacing were calculated around the docking area for all the ligand atom types using AutoGrid4. 100 separate docking calculations were performed for each binder. Each docking calculation consisted of 25×10^6 energy evaluations using the Lamarckian genetic algorithm local search (GALS) method. A low-frequency local search according to the method of Solis and Wets was applied to docking trials to ensure that the final solution represents a local minimum. Each docking run was performed with a population size of 150, and 300 rounds of Solis and Wets local search were applied with a probability of 0.06. A mutation rate of 0.02 and a crossover rate of 0.8 were used to generate new docking trials for subsequent generations. The docking results from each of the 100 calculations were clustered on the basis of root-mean square deviation (RMSD 2.0 Å) between the Cartesian coordinates of the ligand atoms and were ranked on the basis of the free energy of binding. All docking solutions were analysed for the coherency with NMR data and for each compound, the lowest energy solution more in line with the experimental data was further considered and subjected to energy minimisation through the OPLS-2001 force field. All figures were rendered using Chimera software package [16].

2.4. Isothermal titration calorimetry

The $[\text{d}(\text{TGGGGT})_4]$ stock solution was prepared by dissolving the lyophilised compound in 10 mM phosphate buffer with 70 mM KCl, 0.2 mM EDTA, pH 7. The solution was annealed by heating at 95°C for 5 min and slowly cooling to room temperature. The concentration of

the dissolved oligonucleotide was evaluated by UV measurement at 95°C , using as molar extinction coefficient the value calculated by the nearest-neighbour model [17] for the sequence $\text{d}(\text{TGGGGT})$. Stock solutions of the six drugs were prepared by solubilising weighed amounts in DMSO to a final concentration of 8 mM. The complexes between the quadruplex and the drugs were prepared diluting the drug stock solution into the quadruplex solution to get a final DNA:drug molar ratio of 1:4.1 and a final DMSO concentration of 7%. Distamycin A was solubilised in the same buffer used for the oligonucleotide containing 7% of DMSO.

The titrations were carried out in 10 mM phosphate buffer, 70 mM KCl, 0.2 mM EDTA, 7% DMSO, pH 7, at 293 K, using a high-sensitivity ITC-200 microcalorimeter from Microcal (GE Healthcare). In each experiment, volumes of 2 μL of a 1.2 mM distamycin A solution were added into a 50 μL solution of quadruplex–DNA complex, using a computer-controlled 40- μL microsyringe, with a spacing of 180 s between each injection. Each titration was corrected for heat of dilution by subtracting the measured enthalpies of the injections following saturation. Integrated heat data obtained for the titrations were fitted using a non-linear least-squares minimisation algorithm to a theoretical titration curve, using the MicroCal-Origin 7.0 software package from which the binding parameters ΔH° (reaction enthalpy change in kcal mol^{-1}), K_b (binding constant in M^{-1}), and n (stoichiometry) were derived. The entropic contribution was calculated using the relationships $\Delta G^\circ = -RT \cdot \ln K_b$ (R 1.987 $\text{cal mol}^{-1} \text{K}^{-1}$, T 293 K) and $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$.

3. Results and discussion

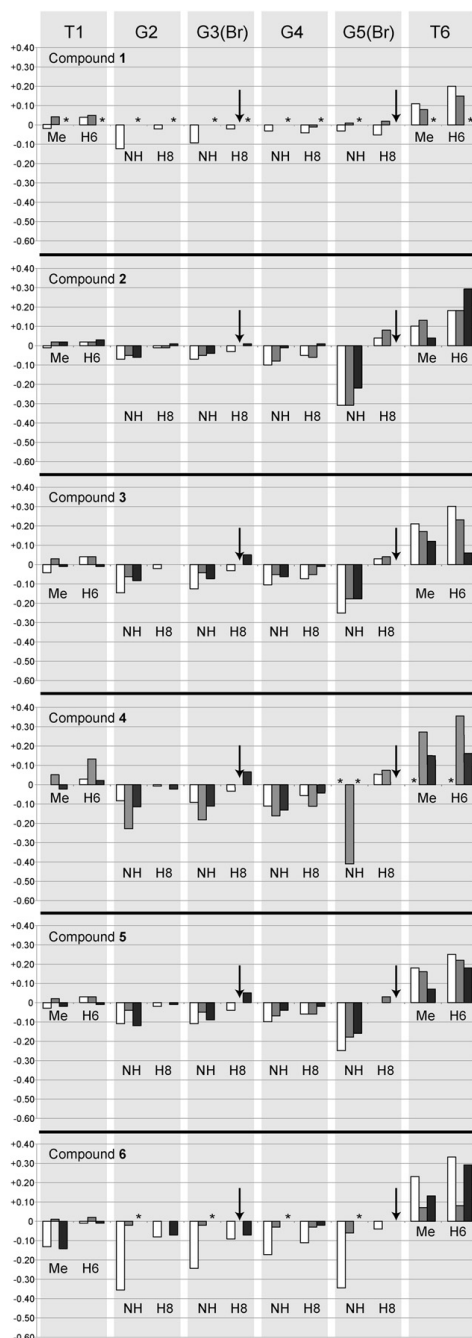
3.1. NMR and molecular modelling studies

Using NMR, six molecules (**1–6**, Fig. 1) were found to be potential groove binders: they cause an appreciable shift, amongst others, of the signals of G3, G4, G5 and T6 of the parallel quadruplex $[\text{d}(\text{TGGGGT})_4]$, indicating that the recognition process involves mostly the 3' side of the grooves, as also computed by the virtual screening calculations [9].

In order to get further insights into the binding mode of compounds **1–6**, we have acquired a number of NOESY experiments of the complexes of the six compounds with the quadruplex $[\text{d}(\text{TGGGGT})_4]$. Unfortunately, as no diagnostic NOE cross-peak could be retrieved for any complex, none of the three-dimensional structures at atomic level could be determined.

Thus, we decided to get a more detailed picture of the interactions from the NMR titration of modified quadruplexes and from molecular docking calculations. As a result, we designed and synthesised modified oligonucleotides, namely $\text{d}(\text{TGG}^{\text{Br}}\text{GGT})$, $\text{d}(\text{TGGG}^{\text{Br}}\text{GT})$ and $\text{d}(\text{TGGGG}^{\text{Br}}\text{T})$, where dG^{Br} is 8-bromo-2'-deoxyguanosine, potentially capable of forming quadruplex structures and possessing a bulky group (bromine) at different positions of the grooves (G3, G4 and G5), as this portion of the DNA should be involved in **1–6** recognition [9]. These modified quadruplexes would then be titrated with the six molecules **1–6** and the results of the NMR titration profile analysed. The rationale behind these experiments derived from the assumption that if these compounds interacted with the quadruplex groove region, the presence of the bromine group should now prevent (or at least limit) the ligand/DNA interactions.

Prior to these experiments, we tested the capability of $\text{d}(\text{TGG}^{\text{Br}}\text{GGT})$, $\text{d}(\text{TGGG}^{\text{Br}}\text{GT})$ and $\text{d}(\text{TGGGG}^{\text{Br}}\text{T})$, to form a quadruplex structure [18]. Their NMR samples were prepared at a concentration of 2 mM, in 0.6 ml ($\text{H}_2\text{O}/\text{D}_2\text{O}$ 9:1) buffer solution having 10 mM KH_2PO_4 , 70 mM KCl, 0.2 mM EDTA, pH 7.0. These samples were then annealed for 5–10 min at 80°C and slowly cooled down to room temperature, then ^1H NMR spectra were recorded by using DPFGSE pulse sequence for H_2O suppression [10,11].



The ^1H NMR spectra (700 MHz, $T = 25^\circ\text{C}$) of $\text{d}(\text{TGGG}^{\text{Br}}\text{GT})$ and $\text{d}(\text{TGGGG}^{\text{Br}}\text{T})$ show the presence of four well defined singlets in the region 11–12 ppm, ascribable to imino protons involved in Hoogsteen hydrogen bonds of G-quartets, as well as the presence of five signals belonging to three guanine H8 and to two thymine H6 protons in the aromatic region. This indicates that a single well defined quadruplex species is present in solution for both molecules, consisting of four G-tetrads and possessing a fourfold symmetry with all strands parallel to each other. In such a quadruplex, each Br group faces the grooves and points outward. On the other hand, the 1D ^1H NMR spectrum of $\text{d}(\text{TGGG}^{\text{Br}}\text{GT})$ shows the presence of a great number of signals in the regions of imino and aromatic protons, suggesting that $\text{d}(\text{TGGG}^{\text{Br}}\text{GT})$ is affected by structural heterogeneity, thus preventing its use in our experiments. Hence, only the quadruplexes $[\text{d}(\text{TGG}^{\text{Br}}\text{GGT})]_4$ and $[\text{d}(\text{TGGGG}^{\text{Br}}\text{T})]_4$ were titrated with compounds 1–6 and the titrations monitored by NMR. A comparison of resonances of protons of the uncomplexed quadruplex and the complexed one has been performed. In particular, we report the $\Delta\delta$ values (chemical shifts of the complex minus free DNA) of aromatic, methyl and imino protons.

As far as compound 1 is concerned, the titration with unmodified quadruplex $[\text{d}(\text{TGGGGT})]_4$ led to a general shift of the monitored signals (white bars in Fig. 2). On the other hand, the titration of $[\text{d}(\text{TGG}^{\text{Br}}\text{GGT})]_4$ causes a slight shift (light-grey bars in Fig. 2) of the residues at the 3' edge of the quadruplex, namely G5-H8, T6-H6/Me, and only of T1-H6/Me at 5' edge, whilst no appreciable shift can be measured for the other signals. This means that the bromine atom at the very centre of the groove did affect the binding of compound 1, and that 1 can be confidently considered a groove binder. Furthermore, the titration of $[\text{d}(\text{TGGGG}^{\text{Br}}\text{T})]_4$ led to a severe line broadening of all signals, making impossible to retrieve any information from the spectra. This NMR phenomenon can be interpreted assuming that the ligand is changing its binding pose on the NMR time scale.

It is noteworthy that compounds 1–6 have been discovered from a previous virtual screening campaign and in that study Autodock4 program was used with a search area large enough to enclose only one of the four identical grooves to avoid redundant information [9]. However, this couldn't cast out the possibility that one or more of the selected molecules can also be able to bind to other part of the target. Thus, herein, we present results of new docking calculations where the search area has been enlarged to comprise the entire surface of the quadruplex $[\text{d}(\text{TGGGGT})]_4$ with the purpose of checking for the capability of the ligands to extend their binding out of the groove. Regarding 1, docking calculations showed that all solutions in the lowest energy families ($\Delta G = -6.2 \text{ kcal mol}^{-1}$), were found to be anchored to the 3' side of the groove in line with the above-mentioned NMR data (1 was predicted in a groove binding mode 75 out of 100 time, see Supporting information). As shown in Fig. 3a (and in Fig. S2a in Supporting information), the amine group of 1 engages a H-bond with the 4'O of the T6 nucleoside, while the ligand carbonyl group together with the *ortho*-hydroxyl group of the phenyl moiety forms H-bonds with G5 base. Three H-bonds have been also detected between 1 and the phosphate backbone. The above-described binding pose has been found 25 out of 100 times (see Supporting information).

Fig. 2. $\Delta\delta$ values of aromatic (H6/H8), methyl (Me) and imino protons (NH) for $[\text{d}(\text{TGGGGT})]_4$ (white bars), $[\text{d}(\text{TGG}^{\text{Br}}\text{GGT})]_4$ (light-grey bars) and $[\text{d}(\text{TGGGG}^{\text{Br}}\text{T})]_4$ (dark-grey bars). Shadowed columns gather resonance variation belonging to the same residue. Asterisks indicate a severe line broadening of the monitored signals. Arrows indicate the lack of bars due to the presence of bromine atoms.

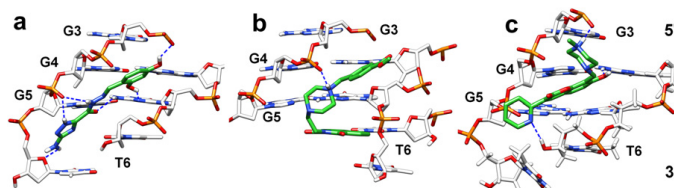


Fig. 3. Binding poses calculated by AD4 for compounds **1** (a), **2** (b) and **3** (c) in the quadruplex structure. DNA backbone is represented as white stick bonds. Ligands are depicted as green sticks. H-bonds are represented as dashed blue lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Surprisingly, compound **2** generates almost identical $\Delta\delta$ values in the titration with $[d(TGGGGT)]_4$ and $[d(TGG^{\text{Br}}GGT)]_4$ (white and light-grey bars, respectively, in Fig. 2), having a general perturbation of the signals belonging to the 3' side of the groove. This can be interpreted by assuming that the bromine atom does not affect the binding of **2** suggesting that this compound should not preferentially bind the very centre of the groove. On the other hand, the titration of $[d(TGGGG^{\text{Br}}T)]_4$ (dark-grey bars in Fig. 2) led to a marked shift of T6-H6/Me signals whereas only a slight shift of all the other monitored signals can be observed. This indicates that **2** prefers to interact with the 3' edge of the quadruplex.

A clearer picture of the ligand–quadruplex interaction could be obtained from docking calculations using again the unmodified quadruplex $[d(TGGGGT)]_4$ as target. The calculations highly converged towards one family of conformations ($\Delta G = -7.4$) (Fig. 3b and Fig. S2b in Supporting information) in which the benzoisoquinolinedione ring stacks on the surface of the 3' terminal quartet, particularly between the T6 and G5 rings, and the rest of the molecule inserts into the groove, where the main anchor point is represented by the protonated nitrogen of the piperazine ring, which establishes a charge reinforced H-bond with the phosphate backbone. This binding mode is in agreement with the new NMR titrations performed in this investigation indicating that **2** is actually characterised by a mixed binding mode, providing both stacking and groove binding interactions. A statistical graph of the binding modes occurrence for compound **2** is reported in Supporting information.

Differently from **1** and **2**, titration of compound **3** with $[d(TGGGGT)]_4$, $[d(TGG^{\text{Br}}GGT)]_4$ and $[d(TGGGG^{\text{Br}}T)]_4$ led to a general shift of almost all signal (Fig. 2). However, the resonances of $[d(TGGGGT)]_4$ (white bars) shifted more and those of $[d(TGG^{\text{Br}}GGT)]_4$ and $[d(TGGGG^{\text{Br}}T)]_4$ less (light- and dark-grey bars, respectively). This indicates that **3** does not possess a well defined binding mode, so that, when the groove is unavailable by the presence of the bromine atoms, **3** slides towards the 3' edge groove of the quadruplex; vice versa, in the case of $[d(TGGGG^{\text{Br}}T)]_4$, the molecule is able to interact only weakly with the available part of the groove. Docking calculations on $[d(TGGGGT)]_4$ suggest that the ligand binds the 3'-end of the groove

($\Delta G = -8.2$) (**3** was predicted in a groove binding mode 87 out of 100 time, see Supporting information). As shown in Fig. 3c (and Fig. S2c in Supporting information), two H-bonds were found, the first between the pyridine nitrogen of **3** and the T6 3'OH group and the second between the hydroxyl group of the chromone core and the G3 NH₂. Hydrophobic contacts between the pyridine and the T6 rings were also detected together with a charge reinforced H-bond established by the protonated nitrogen of the piperazine ring with the phosphate backbone. The above-described binding mode has been found by Autodock program 27 out of 100 time (see Supporting information).

Interestingly, compound **4** displayed different behaviours for the three quadruplexes tested. With the quadruplex $[d(TGG^{\text{Br}}GGT)]_4$ (light-grey bars in Fig. 2), the titration provided higher $\Delta\delta$ values for the residues at the edges of the quadruplex with the respect of unmodified $[d(TGGGGT)]_4$ (white bars in Fig. 2). On the contrary, titration of $[d(TGGGG^{\text{Br}}T)]_4$ caused major shift for the residue sitting in the very centre of the groove (dark-grey bars). This means that compound **4** can interact with the grooves, and, accordingly, when the very centre of the groove is hindered, **4** binds the end sides of the grooves, but it is able to bind the very centre of the grooves when their 3' edges are unavailable. Docking calculations suggest two binding poses especially in line with NMR data (Fig. 4a and b, and Fig. S3a in Supporting information), where the molecule can either interact with 5' residues (binding mode A, $\Delta G = -8.2$) or with the 3' end (binding mode B, $\Delta G = -7.9$). Specifically, in A, the benzoquinone ring establishes a π – π interaction with the T1 ring, while an H-bond between the G2 NH₂ group and the carbonyl moiety is detected. In B, the two carbonyl moieties of the ligand are engaged in H-bonds with G4 NH₂ and T6 NH respectively. In both binding modes a charge reinforced H-bond is observed between the protonated nitrogen of **4** and the phosphate backbone. Based on this finding and on the shift of NMR signals observed for all residues forming the groove, a sliding motion of **4** inside the groove can be proposed. Noteworthy, 100 out of 100 runs propose **4** as groove binder (see Supporting information for further details).

As far as compound **5** is concerned, the titration of $[d(TGGGGT)]_4$ (Fig. 2, white bars) clearly indicates that **5**, like the other selected

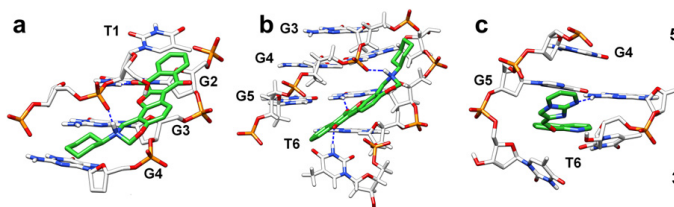


Fig. 4. Two binding poses calculated by AD4 for compound **4** (a and b), and one for compound **5** (c) in the quadruplex structure. DNA backbone is represented as white stick bonds. Ligands are depicted as green sticks. H-bonds are represented as dashed blue lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

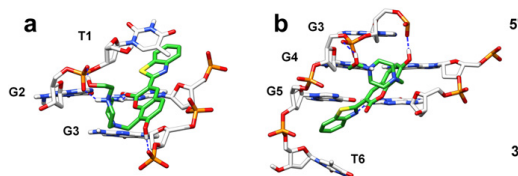


Fig. 5. Two binding poses calculated by AD4 for compound **6**. DNA backbone is represented as white stick bonds. Ligands are depicted as green sticks. H-bonds are represented as dashed blue lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

compounds, prefers to bind the 3' edge of the groove. The titration of $[d(TGG^{\text{Br}}GGT)]_4$ displays significant lower $\Delta\delta$ values for the imino protons (Fig. 2, light-grey bars), while the shifts of the signal of G4, G5 and T6 remain substantially unchanged. On the other hand, the titration of $[d(TGGGG^{\text{Br}}T)]_4$ indicates that **5** is still able to interact with the 3' edge of the quadruplex probably via end-stacking interaction. A mixed binding mode is therefore expected for compound **5**. In line with these data, docking calculations suggested that in **5** the quinolone ring is adapted on the top of the 3'-end sandwiched between the T6 and G5 bases (Fig. 4c and Fig. S3b in Supporting information), with the imidazo-pyrimidine moiety extending towards the groove incipience and H-bonding with the G5 NH_2 group ($\Delta G = -6.7$). The above-described binding mode has been proposed by the docking program 60 out of 100 times (see Supporting information for more details).

Finally, compound **6** displays peculiar titration behaviours. While the titration of the unmodified $[d(TGGGGT)]_4$ showed marked shifts of the G2, G3, G4 and G5 residues (Fig. 2, white bars), the titration of $[d(TGG^{\text{Br}}GGT)]_4$ is instead characterised by very low $\Delta\delta$ values for all residues (light-grey bars). Furthermore, the titration of $[d(TGGGG^{\text{Br}}T)]_4$ mainly caused shifts of the residues G2, G3 and T6, along with a severe line broadening of the imino proton signals (dark-grey bars, Fig. 2). This means that **6** does bind with the grooves of the quadruplex, and, as in **4**, the uniform perturbation of all groove residues of the quadruplex $[d(TGGGGT)]_4$, with respect to the limited ligand size, suggests that a fast sliding motion of **6**

inside the groove is also possible. Molecular docking calculations also revealed that two molecules of **6** could in principle simultaneously anchor the two opposite ends of the groove although the 5' end is preferred (42 times out of 100) (Fig. 5a and b and Fig S3c in Supporting information). Notably, in perfect line with the changes experimentally observed for G2 and G5 aromatic protons signals, the ligand alternatively forms H-bonds with G2 ($\Delta G = -7.5$) or G5 ($\Delta G = -6.1$) bases in the two reported binding modes (Fig. 5). Finally, the two poses suggested by the docking program, in which a π - π interaction is formed with T1 or T6 ring, could somehow account for the T6-H6 proton and T1-methyl signals shift.

3.2. Isothermal titration calorimetry measurements

In order to determine the binding affinity of the compounds **1–6** with the quadruplex $[d(TGGGGT)]_4$, ITC titration experiments were carried out [19].

Previously, we characterised the interaction of distamycin A with the same quadruplex and revealed that four molecules of the drug were accommodated in two opposite grooves of the quadruplex; moreover, the thermodynamic signature of the binding event suggested that the entropic contribution was the one driving the complex formation [8]. An analogous study however was not possible for the six new ligands identified with the virtual screening, because their poor solubility in water prevented us from performing canonical ITC experiments [19], in which a solution of the quadruplex is titrated with a concentrated drug solution in the identical aqueous buffer. As an alternative approach to investigate the affinity of the new drugs for the DNA quadruplex, we carried out competition/displacement experiments, by analysing the ability of distamycin A to bind to the quadruplex in the presence of another compound [20,21]. Despite the solubility concerns, mixtures of the quadruplex and each of the six new drugs were successfully prepared by solubilising the molecules in DMSO and diluting them in a phosphate solution containing the quadruplex (as described in the Methods); these complexes were then titrated with distamycin A, and the outcome followed by ITC.

In Fig. 6 three representative ITC experiments are reported. First, the interaction of distamycin A with $[d(TGGGGT)]_4$ was repeated in

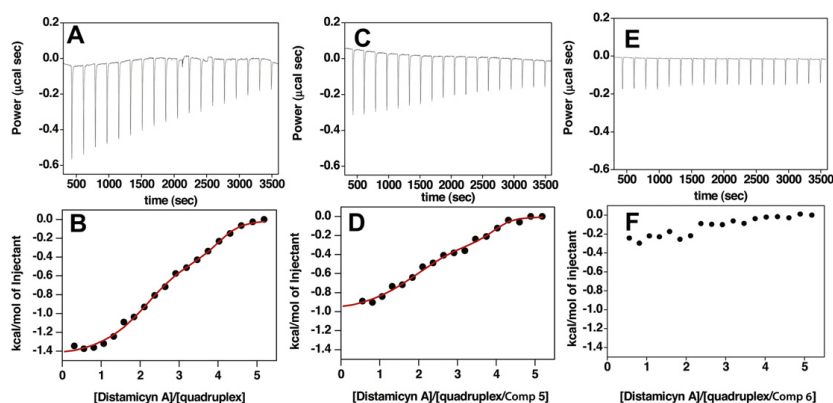


Fig. 6. Calorimetric analysis of the interactions of the quadruplex $[d(TGGGGT)]_4$ and its complexes with the drugs **5** and **6**, with distamycin A. Raw titration data showing the thermal effect of injecting distamycin A into a calorimetric cell containing the quadruplex alone (A), quadruplex with **5** (C) and quadruplex with **6** (E). The normalised heat for the titrations, shown in B, D and F respectively, was obtained by integrating the raw data and subtracting the heat of the ligand dilution. The heat effect reveals the typical double step binding curve of the distamycin A interaction to the quadruplex in B and C, null interaction in F. The red lines in B and D represent the best fit derived by a non-linear least-squares procedure based on an independent binding sites model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1

Thermodynamic parameters for the interaction of Distamycin A with [d(TGGGGT)₄] and its complexes formed with the **1**, **3** and **5**. **2**, **4** and **6** are not reported since they abolish the binding of Distamycin A.

Compound in the calorimetric cell	<i>n</i>	<i>K_D</i> (M ^{−1})	Δ <i>H</i> (kcal mol ^{−1})	−TΔ <i>S</i> (kcal mol ^{−1})	Δ <i>G</i> (kcal mol ^{−1})
[d(TGGGGT) ₄]	1.8	(6.0 ± 0.2) × 10 ⁵	−0.3	−7.4	−7.7
	4.0	(7.0 ± 0.1) × 10 ⁶	−1.5	−7.7	−9.2
[d(TGGGGT) ₄]/ Compound 5	1.9	(3.0 ± 0.1) × 10 ⁵	−0.2	−7.1	−7.3
	4.0	(7.0 ± 0.1) × 10 ⁶	−1.0	−8.2	−9.2
[d(TGGGGT) ₄]/ Compound 3	1.7	(2.0 ± 0.1) × 10 ⁵	−0.6	−6.5	−7.1
	3.8	(7.0 ± 0.1) × 10 ⁶	−1.6	−7.6	−9.2
[d(TGGGGT) ₄]/ Compound 1	1.9	(3.0 ± 0.1) × 10 ⁵	−0.7	−6.6	−7.3
	3.8	(7.0 ± 0.1) × 10 ⁶	−1.6	−7.6	−9.2

The experimental errors on the thermodynamic properties have been evaluated to be <5%.

the new buffer used in this study in absence of any other ligand, clearly showing that in the new experimental conditions this binding event is identical to what was previously observed (Fig. 6, Panels A and B and Table 1). The binding occurs in two separate steps, the first one centred on a stoichiometry of 1:2 in which two molecules of distamycin A interact with the quadruplex and then a second event in which two more molecules bind, leading to a final stoichiometry of 4 molecules of distamycin A per quadruplex. Panels C and E report two displacement experiments carried out titrating with distamycin A the complexes obtained with the drugs **5** and **6** respectively. The two experiments are quite different and indicative of two distinct phenomena. In the first case, a binding curve similar to the control experiment is obtained (compare Fig. 6D and B), indicating that the presence of the drug **5** in the mixture did not have an effect on the association of distamycin A with the [d(TGGGGT)₄]. Distamycin A still binds in two distinct events, with stoichiometry and binding constants not significantly affected by the presence of **5** (Table 1).

Conversely, the ITC titration of distamycin A into the quadruplex/**6** complex (Fig. 6E and F) gives a completely different outcome. In this case, distamycin A appears no longer able to interact with the DNA, suggesting that **6** binds to the quadruplex tightly than distamycin A. It is important to highlight that the competition/displacement experiments do not give any information about the stoichiometry of the complex formed between the quadruplex and the compound competing with the distamycin A. A summary of the results for all six compounds is reported in Table 1: **1** and **3** behave as **5**, whereas the presence of **2** or **4** or **6** abolishes the binding of distamycin A.

In conclusion the competition ITC experiments represent a possible strategy to evaluate whether an insoluble groove binder is stronger or weaker than a soluble drug that could be used as reference point. In our case, using the distamycin A as reference drug, we were able to group the six drugs identified with virtual screening in two extreme classes. The first class contains compounds that bind rather weakly to the quadruplex (**1**, **3** and **5**), whereas in the second class the drugs that bind substantially stronger than distamycin A (**2**, **4** and **6**) are clustered.

4. Conclusions

A successful structure-based virtual screening (VS) campaign was recently undertaken by our research group and it resulted in the identification of six small molecules able to interact with the groove of the quadruplex [d(TGGGGT)₄]. Herein, we report NMR spectroscopy experiments which, combined with extensive molecular docking studies, allow for a more detailed picture of the interaction between each VS-derived binder and the quadruplex DNA. Noteworthy, isothermal titration calorimetry (ITC) measurements revealed that **2**, **4**, and **6**, despite their relatively small dimensions, bind substantially stronger than distamycin A, which is, to the best of

our knowledge, the most potent groove binder identified so far. As it is widely accepted that specificity among the various DNA G-quadruplexes, that might be simultaneously present in the human genome, is a fundamental requirement for the quadruplex binder to become a drug, extensive binding tests towards DNA duplex and different DNA quadruplex topologies will be the next step in our research program. The future results combined with structural studies will provide a source of inspiration for the design of next generation of potent and selective quadruplex DNA drug-like binders.

Acknowledgements

The authors are really grateful to Italian Institute of Technology (IIT) for funding this investigation. MRC is indebted to the Wellcome Trust for the Centre of Biomolecular Spectroscopy at King's.

Appendix. Supporting information

Supporting information associated with this article can be found, in the online version, at doi:10.1016/j.biochi.2011.05.021.

References

- [1] T. de Lange, Shelterin: the protein complex that shapes and safeguards human telomeres, *Genes Dev.* 19 (2005) 2100–2110.
- [2] C. Autauxier, N.F. Lue, The structure and function of telomerase reverse transcriptase, *Annu. Rev. Biochem.* 75 (2006) 493–517.
- [3] W.C. Hahn, S.A. Stewart, M.W. Brooks, S.G. York, E. Eaton, A. Kurachi, R.L. Beijersbergen, J.H. Knoll, M. Meyerson, R.A. Weinberg, Inhibition of telomerase limits the growth of human cancer cells, *Nat. Med.* 5 (1999) 1164–1170.
- [4] a) D. Sun, B. Thompson, B.E. Cathers, M. Salazar, S.M. Kerwin, J.O. Trent, T.C. Jenkins, S. Neidle, L.H. Hurley, Inhibition of human telomerase by a G-quadruplex-interactive compound, *J. Med. Chem.* 40 (1997) 2113–2116; b) D. Ma, C. Che, S. Yan, Platinum(II) complexes with dipyrrophenazine ligands as human telomerase inhibitors and luminescent probes for G-quadruplex DNA, *J. Am. Chem. Soc.* 131 (2009) 1835–1846; c) T. Lai, Discovery of a drug-like G-quadruplex binding ligand by high-throughput docking, *ChemMedChem* 3 (2008) 881–884.
- [5] S.E. Burge, G.N. Parkinson, P. Hazel, A.K. Todd, S. Neidle, Quadruplex DNA: sequence, topology and structure, *Nucleic Acids Res.* 34 (2006) 5402–5415.
- [6] R. Rodriguez, S. Muller, J.A. Yeoman, C. Trentesaux, J.-F. Riou, S. Balasubramanian, A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres, *J. Am. Chem. Soc.* 130 (2008) 15758–15759.
- [7] a) A. de Cian, L. Lacroix, C. Douarre, N. Temime-Smaali, C. Trentesaux, J.-F. Riou, J.-L. Mergny, Targeting telomeres and telomerase, *Biochimie* 90 (2008) 131–155; b) T. Ou, Y. Lu, J. Tan, Z. Huang, K. Wong, L. Gu, G-quadruplexes: targets in anticancer drug design, *ChemMedChem* 3 (2008) 690–713.
- [8] a) A. Randazzo, A. Galeone, L. Mayol, ¹H-NMR study of the interaction of distamycin A and netropsin with the parallel stranded tetraplex [d(TGGGGT)₄], *Chem. Commun.* 11 (2001) 1030–1031; b) A. Randazzo, A. Galeone, V. Esposito, M. Varra, L. Mayol, Interaction of distamycin A and netropsin with quadruplex and duplex structures: a comparative ¹H-NMR study, *Nucleosides Nucleotides Nucleic Acids* 21 (2002) 535–545; c) L. Martino, A. Virno, B. Pagano, A. Virgilio, S. Di Micco, A. Galeone, C. Giancola, G. Bifulco, L. Mayol, A. Randazzo, J. Am. Chem. Soc. 129 (2007) 15950–15956; d) S. Cosconati, L. Marinelli, R. Trotta, A. Virno, S. De Tito, R. Romagnoli, B. Pagano, V. Limongelli, C. Giancola, P.G. Baraldi, L. Mayol, E. Novellino, A. Randazzo, Structural and conformational requisites in DNA quadruplex

- groove binding: another piece to the puzzle, *J. Am. Chem. Soc.* 132 (18) (2010) 6425–6433;
- e) B. Pagano, A. Virno, C.A. Mattia, L. Mayol, A. Randazzo, C. Giancola, Targeting DNA quadruplexes with distamycin A and its derivatives: an ITC and NMR study, *Biochimie* 90 (8) (2008) 1224–1232.
- [9] S. Cosconati, L. Marinelli, R. Trotta, A. Virno, L. Mayol, E. Novellino, A.J. Olson, A. Randazzo, Tandem application of virtual screening and NMR experiments in the discovery of brand new DNA quadruplex groove binders, *J. Am. Chem. Soc.* 131 (2009) 16336–16337.
- [10] T.L. Hwang, A.J. Shaka, Water suppression that works. Excitation sculpting using arbitrary waveforms and pulsed field gradients, *J. Magn. Reson. A* 112 (1995) 275–279.
- [11] C. Dalvit, Efficient multiple-solvent suppression for the study of the interactions of organic solvents with biomolecules, *J. Biomol. NMR* 11 (1998) 437–444.
- [12] J. Jeener, B. Meier, H.P. Bachmann, R.R. Ernst, Investigation of exchange processes by two-dimensional NMR spectroscopy, *J. Chem. Phys.* 71 (1979) 4546–4553.
- [13] D. Marion, M. Ikura, R. Tschudin, A. Bax, Rapid recording of 2D NMR spectra without phase cycling: application to the study of hydrogen exchange in proteins, *J. Magn. Reson.* 85 (1989) 393–399.
- [14] a) R. Huey, G.M. Morris, A.J. Olson, D.S. Goodsell, A semiempirical free energy force field with charge-based desolvation, *J. Comput. Chem.* 28 (2007) 1145–1152;
- b) S. Cosconati, J.A. Hong, E. Novellino, K.S. Carroll, D.S. Goodsell, A.J. Olson, Structure-based virtual screening and biological evaluation of *Mycobacterium tuberculosis* adenosine 5'-phosphosulfate reductase inhibitors, *J. Med. Chem.* 51 (2008) 6627–6630.
- [15] Maestro, Version 9.0.211, Schrodinger, L.L.C., New York, NY, 2009.
- [16] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF chimera – a visualization system for exploratory research and analysis, *J. Comput. Chem.* 25 (13) (2004) 1605–1612.
- [17] C.R. Cantor, R.R. Warshaw, H. Shapiro, Oligonucleotide interactions. 3. Circular dichroism studies of the conformation of deoxyoligonucleotides, *Biopolymers* 9 (1970) 1059–1077.
- [18] a) A. Virgilio, V. Esposito, A. Randazzo, L. Mayol, A. Galeone, 8-Methyl-2'-deoxyguanosine incorporation into parallel DNA quadruplex structures, *Nucleic Acids Res.* 33 (19) (2005) 6188–6195;
- b) L. Petraccone, E. Erra, V. Esposito, A. Randazzo, A. Galeone, G. Barone, C. Giancola, Biophysical properties of quadruple helices of modified human telomeric DNA, *Biopolymers* 77 (2) (2005) 75–85.
- [19] B. Pagano, C.A. Mattia, C. Giancola, Applications of isothermal titration calorimetry in biophysical studies of G-quadruplexes, *Int. J. Mol. Sci.* 10 (2009) 2935–2957.
- [20] S.W. Sigurskjold, Exact analysis of competition ligand binding by displacement isothermal titration calorimetry, *Anal. Biochem.* 277 (2000) 260–266.
- [21] Y.L. Zhang, Z.Y. Zhang, Low-affinity binding determined by titration calorimetry using a high-affinity coupling ligand: a thermodynamic study of ligand binding to protein tyrosine phosphatase 1B, *Anal. Biochem.* 261 (1998) 139–148.

Shooting for Selective Druglike G-Quadruplex Binders: Evidence for Telomeric DNA Damage and Tumor Cell Death

Sandro Cosconati,^{†,‡} Angela Rizzo,^{‡,‡} Roberta Trotta,[§] Bruno Pagano,[§] Sara Iachettini,[‡] Stefano De Tito,[§] Ilaria Lauri,[§] Iolanda Fotticchia,^{||} Mariateresa Giustiniano,[§] Luciana Marinelli,[§] Concetta Giancola,^{||} Ettore Novellino,[§] Annamaria Biroccio,^{*,‡} and Antonio Randazzo^{*,§}

[†]Dipartimento di Scienze Ambientali, Seconda Università di Napoli, 81100 Caserta, Italy

[‡]Experimental Chemotherapy Laboratory, Regina Elena National Cancer Institute, 00158 Rome, Italy

[§]Dipartimento di Chimica Farmaceutica e Tossicologica, Università di Napoli "Federico II", 80131 Napoli, Italy

^{||}Dipartimento di Scienze Chimiche, Università di Napoli "Federico II", 80126, Naples, Italy

Supporting Information

ABSTRACT: Targeting of DNA secondary structures, such as G-quadruplexes, is now considered an appealing opportunity for drug intervention in anticancer therapy. So far, efforts made in the discovery of chemotypes able to target G-quadruplexes mainly succeeded in the identification of a number of polyaromatic compounds featuring end-stacking binding properties. Against this general trend, we were persuaded that the G-quadruplex grooves can recognize molecular entities with better drug-like and selectivity properties. From this idea, a set of small molecules was identified and the structural features responsible for G-quadruplex recognition were delineated. These compounds were demonstrated to have enhanced affinity and selectivity for the G-quadruplex over the duplex structure. Their ability to induce selective DNA damage at telomeric level and to induction of apoptosis and senescence on tumor cells is herein experimentally proven.



INTRODUCTION

Telomeres are critical chromosomal elements which guarantee proper replication and protection of chromosome ends.¹ They are made up of 2–20 kb of double-stranded TTAGGG repeats and feature a 3' single-stranded overhang of 50–500 nucleotides. Parallel to normal cells proliferation, telomeres get gradually shorter triggering irreversible growth arrest (cellular senescence).² A telomere maintenance mechanism is provided by the six-membered protein complex called shelterin and by telomerase. The latter adds copies of the repeated motif to the end of the single-stranded overhang. This enzyme is transcriptionally repressed in most differentiated human somatic cells³ while being overexpressed in about 85% of cancer cells.⁴ In the remaining 15% of human tumors, telomere lengthening is obtained by a different mechanism known as alternative lengthening of telomere (ALT).⁵ In both cases, telomeres are maintained to a stable length with consequent senescence circumvention and cellular immortalization.⁶ In this scenario, it is now widely accepted that telomere maintenance and protection play a central role in tumorigenesis. Thus, agents that are able, at any level, to influence telomere homeostasis are considered now an appealing opportunity for drug intervention in anticancer therapy. The 3' single-stranded overhang of the telomeric DNA in eukaryotic cells can adopt the peculiar G-quadruplex fold.⁷ The stabilization of this fold through the interaction with different ligands alters the G-rich overhang structure and causes its degradation through a DNA-

damage repair pathway and release of one of shelterin proteins (i.e., POT1) from telomeres.⁸ These events lead to a fast induction of tumor cell senescence and apoptosis. For some of these compounds the *in vivo* anticancer activity has been reported (BRACO-19, RHSP4, and telomestatin).^{9–11} Unfortunately, none of these molecules have progressed beyond the experimental stage into clinical trial, mainly because of insufficient druglike properties. Recently, our pursuit of new G-quadruplex ligands succeeded in the discovery of six leadlike chemotypes that were proven to effectively interact with the [d(TGGGGT)]₄ G-quadruplex structure.¹² Among them, compound **1a** (Table 1) appeared to be the most promising hit. So far, ligands that selectively recognize G-quadruplex grooves are few,¹³ even though quadruplex groove recognition is likely to provide much more quadruplex-selective ligands. Interestingly, **1a** was proven to span the entire quadruplex grooves and was demonstrated to interact more tightly than distamycin A (Dst),¹⁴ which was described as the most affine G-quadruplex groove binder.¹⁵ More general considerations need to be done on the druglike properties of the **1a** coumarin core. This is a naturally occurring structure (mainly in plants) that is present in a plethora of compounds endowed with different biological activities (anti-HIV, CNS-active, anticoagulant, anti-inflammatory, antitumor).¹⁶ Because of the

Received: July 12, 2012

Published: October 11, 2012



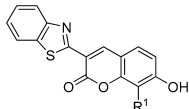
ACS Publications

© 2012 American Chemical Society

9785

dx.doi.org/10.1021/jm301019w | J. Med. Chem. 2012, 55, 9785–9792

Table 1. Structures of the Selected Coumarin Derivatives



Cpd	R ¹	Cpd	R ¹
1a		1d	
1b		1e	
1c		1f	

number of documented biological activities and its amenability to combinatorial chemistry, the coumarin scaffold represents a well-known example of privileged structure.¹⁶

In the present study, the promising derivative **1a** was used as a seed for searching similar entities in several commercially available databases, and NMR experiments allowed identification of a small focused library of structural analogues with G-quadruplex binding properties. By a back and forth approach, the structural features responsible for G-quadruplex groove recognition were delineated, while isothermal titration calorimetry (ITC) measurements allowed for the identification of chemotypes featuring a tighter binding than Dst. Different from Dst, the best binders were also proved to be G-quadruplex selective over duplex. These results propelled the biological characterization of the new ligands, demonstrating their ability to induce selective DNA damage at telomeric level and induction of apoptosis and senescence on tumor cells.

RESULTS AND DISCUSSION

Compound Selection and G-Quadruplex NMR Binding Assays. Given the relative synthetic accessibility of coumarin compounds, we first decided to search in commercial molecular databases if analogues of compound **1a** were available. This approach has the advantage of rapidly providing a library of structural analogues of the lead compound and also being highly economically efficient. Therefore, the simplistic pairwise Tanimoto similarity score was computed between **1a** and the compounds present in the full ZINC database collection of purchasable compounds (~7 million compounds) using a 70% similarity threshold. This resulted in 272 compounds that were visually analyzed revealing that, as expected, coumarin derivatives were selected (Table 1) as well as several analogues featuring the regioisomeric chromone scaffold (Table 2).

With the aim of exploring the influence of the pendant amine moiety on the quadruplex binding, five of the available coumarin derivatives derived from the filtered database were

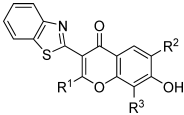
first purchased (**1b–f**). These were all tested for their ability to interact with the DNA quadruplex through nuclear magnetic resonance (NMR) spectroscopy, which is a valuable tool for studying molecular interactions in solution.¹⁷ A number of DNA quadruplex folding topologies are known, and the kind of folding is strongly affected by the sequence, buffer, and molecular crowding conditions. This is particularly true for the human telomeric sequence.⁷ In fact, several structures have been reported having different strand orientations and loop distributions. Unfortunately, the topology of the quadruplex structure adopted by the human telomeric sequence *in vivo* is unknown, so in analogy to what was done in our previous investigations,^{12,15} we have decided to use the highly symmetric tetramolecular quadruplex [d(TGGGGT)]₄. In particular, we employed the chemical shift perturbation method¹⁸ to detect interaction between a molecular candidate and the G-quadruplex structure. The signals that can be most easily monitored are the imino and aromatic protons of all bases and the methyl protons of thymines, as they all are in the less crowded region of the NMR spectrum of [d(TGGGGT)]₄.




















Since we were mainly interested in potential groove binder analogues, in our inspection we considered quadruplex ligands all compounds providing a shift of the G3 and G4 signals of at least 0.05 ppm. Following this criterion, among the five tested coumarins, compounds **1b–d** demonstrated appreciable G-quadruplex binding while **1e** and **1f** displayed no DNA interaction. This indicates that the presence a H-bond donor (**1b**) and/or a positively charged group at a distal position (**1c–d**) with respect to the coumarin scaffold is critical for DNA binding.

More structural variability was present when considering the regioisomeric chromone derivatives (**2a–k**, Table 2). This moiety is also regarded as a privileged structure, being present in different biologically active compounds and prone to combinatorial synthesis.¹⁹ In this case, we first tested the G-quadruplex binding properties of **2a** through NMR titration. This experiment did not indicate appreciable shift of any signal of [d(TGGGGT)]₄, suggesting that the chromone scaffold is unable to provide efficient binding for the quadruplex. On the other hand, the positively charged group in position 8 (R³ substituent) should provide this moiety with a critical interaction point with the quadruplex (most probably, with the phosphate backbone atoms), as confirmed by the presence of detectable interactions of compounds **2b** and **2c** with the target. Furthermore, bulky tertiary amines are not tolerated (**2d**), while cyclic amines (**2e–k**) are able to cause substantial shift of the G3, G4, G5, and T6 signals. In order to probe the influence of a substituent in position 2 on the chromone scaffold (R¹ in Table 2), **2l–p**, featuring H-bond donor groups, were also tested. Interestingly, while **2l–o** still provide appreciable quadruplex binding, **2p** does not interact with the DNA, suggesting that bulkier amines are detrimental for the binding regardless of the presence of a H-bond donor in position 2. Indeed, the latter position seems to directly influence the G-quadruplex recognition, considering that the simple substitution with a methyl group (**2s–t**) abolishes the ligand binding as demonstrated by NMR experiments. The same holds true for position 6 (R²) of the chromone ring that when substituted with a ethyl chain results in compounds **2q–r**, which are incapable of [d(TGGGGT)]₄ binding.

Isothermal Titration Calorimetry (ITC) Experiments. For compounds demonstrating appreciable interaction with the DNA G-quadruplex (**1a,b**, **2b,c**, **2e–o**) displacement iso-

Table 2. Structures of the Selected Chromone Derivatives



Cpd	R ¹	R ²	R ³	Cpd	R ¹	R ²	R ³
2a	-H	-H	-H	2k	-H	-H	
2b	-H	-H		2l	-NH ₂	-H	
2c	-H	-H		2m	-NH ₂	-H	
2d	-H	-H		2n	-NH ₂	-H	
2e	-H	-H		2o	-NH ₂	-H	
2f	-H	-H		2p	-NH ₂	-H	
2g	-H	-H		2q	-H	-CH ₂ CH ₃	
2h	-H	-H		2r	-H	-CH ₂ CH ₃	
2i	-H	-H		2s	-CH ₃	-H	
2j	-H	-H		2t	-CH ₃	-H	

thermal titration calorimetry (ITC) experiments²⁰ were performed to investigate their quadruplex binding affinity. Since it is well-known that multiple conformations of the human telomeric sequences may coexist in solution and that such polymorphism could invalidate the data obtained by ITC measurements, we have decided to use, in analogy to NMR, the [d(TGGGGT)]₄ quadruplex.

Unfortunately, efforts to obtain direct thermodynamic information from canonical ITC experiments failed because of solubility issues regarding the ligands at the rather high concentrations required for such measurements.²¹ Displacement experiments were then effectively carried out by analyzing the binding of Dst to the G-quadruplex previously saturated with each ligand (see Experimental Section). Although the displacement ITC experiments do not allow direct measurement of the thermodynamic parameters involved in the binding processes, this strategy represents a valid approach for evaluating the efficiency of a G-quadruplex (or duplex) binder

compared to Dst. Indeed, when the DNA–ligand complex is formed, the ligand will inhibit the binding of Dst if its affinity for the G-quadruplex is higher than the latter; conversely, it will be displaced by a stronger binder.

Figure 1a and Figure 1b show two representative ITC displacement experiments carried out by titrating with Dst the [d(TGGGGT)]₄ saturated with compounds **2o** and **2l**, respectively. The results obtained in the two experiments clearly show two different phenomena. In the first case, the ITC profile for the titration of ligand-saturated G-quadruplex is essentially identical to the one obtained for the binding of Dst to the G-quadruplex alone,¹⁴ thus indicating that the presence of **2o** does not affect significantly the interaction. On the contrary, the titration of [d(TGGGGT)]₄/2l mixture with Dst (Figure 1b) gives completely different results. In this case, ITC data show constant heat release at each injection of Dst, only due to ligand dilution, proving that it is no longer able to interact with the G-quadruplex. Overall, ITC experiments

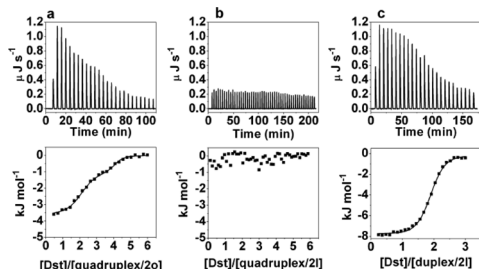


Figure 1. Raw ITC data (top panels) and integrated heat (bottom panels) for titration of $[d(TGGGGT)]_4/2o$ (a), $[d(TGGGGT)]_4/2l$ (b), and $d(CGCGAATTCGCG)_2/2l$ (c) mixtures with Dst at 25 °C. The integrated heat for the titrations (squares) was obtained by integrating the raw data and subtracting the heat of the ligand dilution. The lines represent nonlinear least-squares fit of the data to the appropriate binding model.

showed that 4 out of 15 compounds (**1a**, **2c**, **2l**, and **2m**) have the ability to bind the G-quadruplex more tightly than Dst, inhibiting its interaction. To evaluate a possible selectivity of the best ligands for G-quadruplex over duplex, we performed identical ITC displacement experiments by titrating $d(CGCGAATTCGCG)_2$ duplex/ligand mixtures with Dst. This self-complementary DNA dodecamer was chosen because it contains the central AATT core, considered being one of the specific binding sites for Dst.²² Figure 1c shows an example of raw ITC data and binding isotherm for the titration of $d(CGCGAATTCGCG)_2/2l$ mixture with Dst. As shown, at each injection of Dst solution, less and less heat release was measured until constant values were obtained, implying a saturable process. The binding isotherm shows a typical sigmoidal binding curve and clearly suggests that the presence of the compound in the mixture has no effect on the interaction of Dst with the duplex. Similar results were obtained for **2c** and **2m** (see Supporting Information), suggesting G-quadruplex selectivity of these three compounds. On the other hand, **1a** was shown to affect Dst-duplex interaction, thus suggesting a poor selectivity.

The Newly Identified G-Quadruplex Ligands Induce DNA Damage and Cell-Cycle Arrest. These encouraging results propelled the full biological characterization of the new ligands for investigating the ability of the new ligands to cause telomere uncapping (Figure 2). To this aim, a two-step analysis was performed to establish, in the first one, if the compounds were able to induce DNA damage and, in the second one, if the DNA damage was localized to the telomeres. By using human transformed BJ fibroblasts (BJ-HELT), we found that, different from Dst, all the ligands were able to induce DNA damage (at least at the higher drug dose, Figure 2a) and deconvolution microscopy analysis showed that some of the damaged foci colocalized with TRF1, a good marker for interphase telomeres²³ forming the so-called TIF (telomere dysfunction induced foci)²⁴ (Figure 2d). Of note, quantitative analysis identified compounds **2c**, **2l**, and **2m** as the most potent in inducing telomere damage: the percentage of cells with more than four γ H2AX/TRF1 colocalizations reached about 50% (Figure 2b), with a mean of about eight TIF per nucleus (Figure 2c). Such a biological evaluation also confirmed that the sole chromone structure (**2a**) was not proved to cause

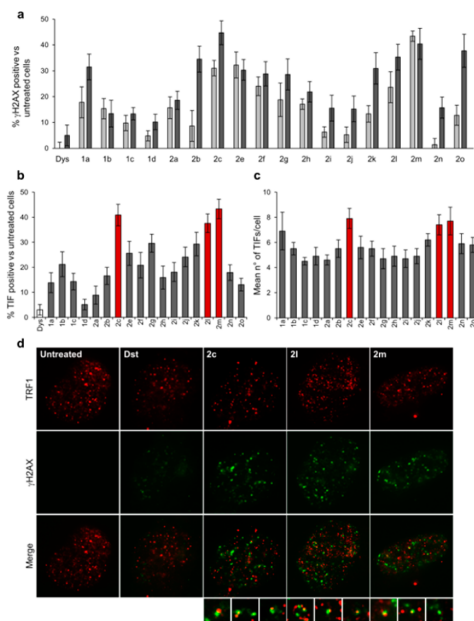


Figure 2. DNA damage activation at telomeres. BJ-HELT fibroblasts were treated for 24 h with Dst and the indicated ligands at doses 0.1 (light-gray bars) and 0.5 μ M (dark-gray bars). Cells were processed for immunofluorescence (IF) using antibodies against γ H2AX and TRF1 to mark DNA damage and telomeres, respectively. Percentages of γ H2AX-positive (a) and TIF-positive (b) treated vs untreated cells are reported in the histograms. (c) Mean number of TIF in the indicated samples. Cells with four or more γ H2AX/TRF1 foci were scored as TIF positive. The red bars highlight the most effective ligands. Error bars indicate the standard deviation. (d) Representative images of IF of untreated and Dst-, **2c**-, **2l**-, and **2m**-treated BJ-HELT cells. Enlarged views of TIFs are reported below the merged images. The images were acquired with a Leica deconvolution microscope (magnification 100 \times).

substantial DNA damage. Interestingly, the whole data are in perfect consonance with ITC experiments that indicated that **2c**, **2l**, and **2m** are strong, selective G-quadruplex binders. A further correlation can also be done for compound **1a** that, by promiscuously binding the DNA (see ITC data), is able to induce high levels of DNA damage but few TIFs.

The above results raised the interesting possibility that telomere damages induced by the ligands in transformed fibroblasts may rapidly and efficiently promote growth inhibition in tumor cells. Treatment of HeLa cells with one of the most promising selected ligands (**2l**) triggered a dose-dependent inhibition of cell survival (Figure 3a) associated with an early accumulation of cells in the G₂/M phase of the cell cycle, and at 96 h of treatment a fraction of cell population resided in the sub-G₁ compartment, indicative of apoptosis (Figure 2b).

Apoptosis induction triggered by **2l** has been confirmed by annexin staining (Figure 3c; at 96 h of treatment about 30% of cells are annexin V-positive/PI negative), and it was also accompanied by the induction of a senescence phenotype: large

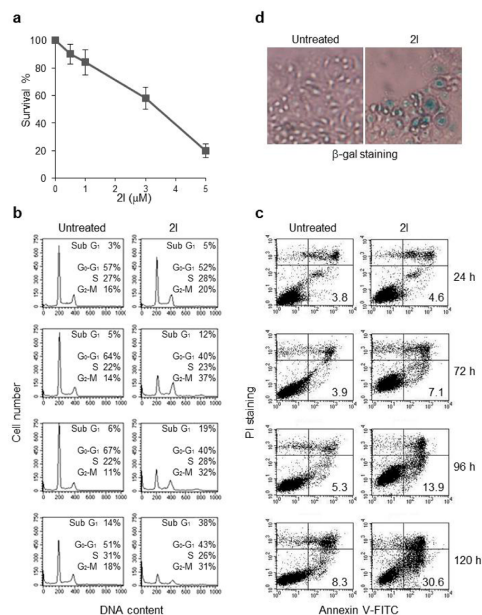


Figure 3. Biological effects of 2l ligand. (a) Survival curve of the HeLa cells exposed to different doses of 2l ranging from 0.5 to 5 μM. (b) Cell cycle analysis and (c) apoptosis evaluation of HeLa cells processed at the indicated times after exposure with 3.5 μM 2l. (c) Biparametric dot plots showing PI vs annexin V staining in the indicated samples. (d) SA-β-gal staining of HeLa cells untreated and treated with 5 μM 2l for 5 days.

cell size, vacuolated cytoplasm, and β-galactosidase activity (Figure 3d).

The Newly Identified G-Quadruplex Ligands Feature Enhanced Druglikeness. In our inspection, the most interesting ligands (1a, 2c, 2l, and 2m) were also compared for their predicted absorption, distribution, metabolism, and excretion (ADME) properties with respect to other compounds previously described as G-quadruplex binders (Supporting Information, Table S3). These calculations were performed employing the QikProp software (QikProp, version 3.4 (2011); Schrödinger, LLC, New York, NY). In addition to predicting molecular properties, QikProp provides ranges for comparing each compound's property with those of 95% of known drugs (these ranges are provided in Supporting Information, Table S4). This software was also used because it allows for flagging reactive functional groups that may cause false positives in biological assays. Results of QikProp calculations are reported in Supporting Information, Tables S5–S7. According to these calculations, only 7 out of 32 inspected compounds display no violations of the ranges recommended for each descriptor or property. Strikingly, compounds 1a, 2c, 2l, and 2m are among these most promising ligands.

CONCLUSIONS

Targeting of DNA secondary structures such as G-quadruplexes is now considered an appealing opportunity for drug

intervention in anticancer therapy.²⁵ So far, efforts made in the discovery of chemotypes able to target G-quadruplexes mainly succeeded in the identification of a number of polyaromatic compounds featuring end-stacking binding properties. Unfortunately, the poor druglike properties of these compounds turned out to be a main limitation during the in vivo verification of their antitumor properties. Herein, with the aim of discovering G-quadruplex groove binders with enhanced druglike properties, a lead optimization campaign was undertaken starting from a promising virtual screening hit. Thus, the physicochemical characterization (NMR and ITC) of the binding of a set of closely related analogues allowed identification of novel ligands of the [d(TGGGGT)]₄ quadruplex. Interestingly, their biological characterization demonstrated the ability to induce selective DNA damage at telomeric level and induction of apoptosis and senescence on tumor cells. These results substantiate our choice of using the [d(TGGGGT)]₄ structure as a working model to design new molecular entities endowed with G-quadruplex binding properties. Furthermore, for the first time, we demonstrate that selective G-quadruplex binding and telomeric DNA damage can be elicited by more druglike chemotypes. These findings pave the way for the design of new potential drugs and shed new insights into the emerging field of DNA quadruplex.

EXPERIMENTAL SECTION

Oligonucleotide Synthesis. The oligonucleotide d(TGGGGT) was synthesized using standard protocol.²⁶ The oligomer was detached from the support and deprotected by treatment with concentrated aqueous ammonia at 55 °C for 12 h. The combined filtrates and washings were concentrated under reduced pressure, redissolved in H₂O, analyzed, and purified by high-performance liquid chromatography (HPLC) on a Nucleogel SAX column (Macherey-Nagel, 1000-8/46), using buffer A consisting of 20 mM KH₂PO₄/K₂HPO₄ aqueous solution (pH 7.0), containing 20% (v/v) CH₃CN, buffer B consisting of 1 M KCl, 20 mM KH₂PO₄/K₂HPO₄ aqueous solution (pH 7.0), containing 20% (v/v) CH₃CN, a linear gradient from 0% to 100% B for 30 min, and flow rate 1 mL/min. The fractions of the oligomer were collected and successively desalted by Sep-pak cartridges (C-18). The isolated oligomer proved to be >99% pure by NMR.

Selected Compounds. The selected compounds were purchased from the supplier as indicated in Table S1 (see Supporting Information). The purity of compounds 1a–f and 2a–t was assessed using reversed-phase high-performance liquid chromatography (HPLC), using a Shimadzu C18, 5 μm (150 mm × 4.6 mm) column. The elution was performed with a 1.0 mL/min flow rate using a linear gradient from 0% to 100% methanol in water over 30 min. The detection was performed at 210 nm. The purity was also analyzed with high-performance liquid chromatography–mass spectrometry (HPLC–MS) performed on an Agilent 1200 series system (Agilent Technologies, Santa Clara, CA, USA) equipped with an Agilent 6110 series LC/MS quadrupole instrument, using a Phenomenex Luna C18, 5 μm (150 mm × 4.6 mm) column. The elution was performed with a 1.0 mL/min flow rate using a linear gradient from 0% to 90% acetonitrile in water over 20 min. Detection was performed at 210 nm. The relative purity of compounds 1a, 1c–e, 2a–d, 2f, 2h, 2i, 2l–n, and 2p–r was higher than 98.0%. Purity of compounds 1b, 1f, 2e, 2g, 2j, 2k, 2o, 2s, and 2t ranged between 95% and 98%.

Nuclear Magnetic Resonance Experiments. The quadruplex NMR samples were prepared at 0.1 mM (0.4 mM single strand concentration) in 0.2 mL (H₂O/D₂O, 9:1) of buffer solution having 10 mM KH₂PO₄, 70 mM KCl, 0.2 mM EDTA, pH 7.0. NMR spectra were recorded with Varian ¹H INOVA 700 MHz spectrometer. ¹H chemical shifts were referenced to external sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS). 1D proton spectra of the sample in H₂O were recorded using pulsed-field gradient DFGSE^{27,28} for H₂O

suppression. The NMR data were processed on an iMAC running iNMR software (www.inmr.net).

Chemical Shift Perturbation Experiments. The quadruplex [d(TGGGGT)]₄ has been titrated with each of the selected compounds. The samples **1e–f**, **2a**, **2d**, **2p–t** were not able to cause any significant shift of the DNA signals. On the other hand, the compounds **1b–d**, **2b**, **2c**, **2e–k**, **2l–o** turned out to be able to bind the quadruplex. Particularly, for all the titrations of **1b–d**, **2b**, **2c**, **2e–k**, **2l–o**, the four DNA strands turned out to be magnetically equivalent throughout the titration, and no splitting of resonances was observed at any stage. In order to preliminarily evaluate the binding site of each analogue, a comparison of resonances of some protons of the uncomplexed DNA and the complexed one has been done. In particular, we report the $\Delta\delta$ values (chemical shifts of the complex minus free DNA) of aromatic, methyl, and imino protons (Supporting Information, Table S2). Generally, the signals of the protons of the T1 residue shifted the least, whereas the ones of residue T6 shifted more. In any case, a general shift of the aromatic and imino signals was also observed for the G2, G3, G4, and G5.

Isothermal Titration Calorimetry. The d(TGGGGT) and d(CGCGAATTCGCG) oligonucleotide sequences were prepared by dissolving the lyophilized compound in 10 mM phosphate buffer with 70 mM KCl, 0.2 mM EDTA, pH 7. The solutions were annealed by heating at 90 °C for 5 min and slowly cooling to room temperature and then equilibrated at 4 °C for 24 h. The concentration of oligonucleotides was determined by UV adsorption measurements at 90 °C using molar extinction coefficient values $\epsilon_{(260\text{ nm})}$ of 57 800 and 110 700 M⁻¹ cm⁻¹ for d(TGGGGT) and d(CGCGAATTCGCG), respectively. The molar extinction coefficients were calculated by the nearest neighbor model.²⁹ Stock solutions of the investigated compounds were prepared by solubilizing weighted amounts in DMSO to a final concentration of 8 mM. The mixtures of the DNA molecules and the compounds were prepared by diluting the ligand stock solution into the DNA solution to get a final ligand/DNA molar ratio of 4:1 and a final DMSO concentration of 7%. Dst was solubilized in the same buffer used for the mixtures. ITC experiments were performed at 298 K using a CSC 5300 Nano-ITC microcalorimeter from Calorimetry Science Inc. (Lindon, UT) with a cell volume of 1 mL. The titrations were carried out in 10 mM phosphate buffer, 70 mM KCl, 0.2 mM EDTA, 7% DMSO, pH 7. In each experiment, volumes of 5–10 μ L of Dst solution (360–720 μ M) were added into a 30 μ M solution of DNA or DNA/ligand mixture, using a computer-controlled 250 μ L microsyringe, with a spacing of 200–400 s between each injection. Each titration was corrected by subtracting the heat of Dst dilution. Where possible, integrated heat data obtained for the titrations were fitted by employing a nonlinear least-squares minimization algorithm to a theoretical titration curve, using the Bindwork software from Calorimetry Science Inc.

Cells and Culture Conditions. Transformed human BJ fibroblasts expressing hTERT and SV40 early region (BJ-HELT⁺) and human epithelial carcinoma cell line (HeLa) were obtained as previously reported⁸ and grown in Dulbecco modified Eagle medium (D-MEM, Invitrogen Carlsbad, CA, U.S.) supplemented with 10% fetal calf serum, 2 mM L-glutamin, and antibiotics.

Immunofluorescence. Immunofluorescence was performed as previously reported.³⁰ Cells were fixed in 2% formaldehyde and permeabilized in 0.25% Triton-X100 in PBS for 5 min at room temperature. For immunolabeling experiments, cells were incubated with primary antibody, then washed in PBS and incubated with the secondary antibodies. The following primary antibodies were used: pAb anti-TRF1 (Abcam Ltd.; Cambridge, U.K.); mAb anti- γ H2AX (Upstate; Lake Placid, NY). The following secondary antibodies were used: TRITC conjugated goat anti-rabbit, FITC conjugated goat anti-mouse (The Jackson Laboratory). Fluorescence signals were recorded by using a Leica DMIRE2 microscope equipped with a Leica DFC 350FX camera and elaborated by a Leica FW4000 deconvolution software (Leica, Solms, Germany).

Clonogenic Assay. HeLa cells were seeded in 60 mm Petri dishes (Nunc, MasciaBrunelli, Milan, Italy) at a density of 5×10^2 cells per dish and 24 h later exposed to different doses (ranging from 0.5 to 5

mM) of **2m**. Cell colony-forming ability was determined as previously described.³⁰ All the experiments were repeated four times in triplicate.

Flow Cytometric Analysis. The cell cycle analysis was performed by flow cytometry. Cells were washed in PBS and fixed in 70% ethanol in PBS. 1×10^6 cells were centrifuged and resuspended in a staining solution (50 μ g/mL PI, 75 kU/mL RNase A in PBS) for 30 min at room temperature in the dark and analyzed by flow cytometry using FACSscalibur (Becton-Dickinson, San Jose, CA, U.S.). For each analysis 20 000 events were collected. Cell cycle distribution and percentage of apoptotic cells were analyzed using Cell Quest (BDIS) and ModFit LT (Verity Software House, Topsham, ME).

Evaluation of Apoptosis. Apoptosis was detected by flow cytometric analysis of annexin V staining. Annexin V-FITC vs PI assay (Vibrant apoptosis assay, V-13242, Molecular Probes, Eugene, OR, U.S.) was performed as previously described.³¹ Briefly, adherent cells were harvested and suspended in the annexin-binding buffer (1×10^6 cells/mL). Thereafter, cells were incubated with annexin V-FITC and PI for 15 min at room temperature in the dark and immediately analyzed by flow cytometry. The data are presented as biparametric dot plots showing PI red fluorescence vs annexin V-FITC green fluorescence.

Senescence Analysis. Senescence-associated β -galactosidase (SA- β -gal) staining on HeLa cells was performed as described by Dimri et al.³² Briefly, after exposure with 3.5 μ M **2m** for 5 days to cell culture, the cells were fixed with 2% glutaraldehyde in PBS for 5 min at room temperature, washed in PBS, and incubated at 37 °C for 24 h in staining solution: 1 mg/mL 5-bromo-4-chloro-3-indolyl- β -D-galactoside (X-gal), 5 mM potassium ferrocyanide, 5 mM potassium ferricyanide, 2 mM MgCl₂ in PBS, pH 6.0. Then cells were analyzed using an optical microscope.

Statistical Analysis. The experiments have been repeated from three to five times, and the results obtained are presented as the mean \pm SD. Significant changes were assessed by using Student's *t* test for unpaired data, and *P* < 0.05 was considered significant.

■ ASSOCIATED CONTENT

● Supporting Information

Vendor codes and suppliers for each tested compound; differences in chemical shifts ($\Delta\delta$) of selected signals of DNA upon binding of derivatives **1b–d** and **2b,c,e–k,l–o**; additional isothermal titration calorimetry experiments; results of QikProp calculations. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*For A.B.: phone, +39-06-52662569; e-mail, biroccio@ifo.it. For A.R.: phone, +39-081-678514; e-mail, antonio.randazzo@unina.it.

Author Contributions

[†]These authors contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Thanks are due to Carmen D'Angelo for flow cytometric analysis. This work was supported by the Italian Institute of Technology (IIT), Italian Association for Cancer Research (A.I.R.C. No. 11567 and No. 11947), and Italian MIUR (PRIN 2009). S.I. is recipient of fellowships from Italian Foundation for Cancer Research (FIRC).

■ ABBREVIATIONS USED

NMR, nuclear magnetic resonance; ITC, isothermal titration calorimetry; ALT, alternative lengthening of telomere; TIF, telomere dysfunction induced foci

■ REFERENCES

- (1) de Lange, T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* **2005**, *19*, 2100–2110.
- (2) Blackburn, E. H. Switching and signaling at the telomere. *Cell* **2001**, *106*, 661–673.
- (3) Lin, S. Y.; Elledge, S. J. Multiple tumor suppressor pathways negatively regulate telomerase. *Cell* **2003**, *113*, 881–889.
- (4) Kim, N. W.; Piatyszek, M. A.; Prowse, K. R.; Harley, C. B.; West, M. D.; Ho, P. L.; Coviello, G. M.; Wright, W. E.; Weinrich, S. L.; Shay, J. W. Specific association of human telomerase activity with immortal cells and cancer. *Science* **1994**, *266*, 2011–2015.
- (5) Cesare, A. J.; Reddel, R. R. Telomere uncapping and alternative lengthening of telomeres. *Mech. Ageing Dev.* **2008**, *129*, 99–108.
- (6) Bodnar, A. G.; Ouellette, M.; Frolkis, M.; Holt, S. E.; Chiu, C. P.; Morin, G. B.; Harley, C. B.; Shay, J. W.; Lichtsteiner, S.; Wright, W. E. Extension of life-span by introduction of telomerase into normal human cells. *Science* **1998**, *279*, 349–352.
- (7) (a) Chaires, J. B. Human telomeric G-quadruplex: thermodynamic and kinetic studies of telomeric quadruplex stability. *FEBS J.* **2010**, *277*, 1098–1106. (b) Phan, A. T. Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J.* **2010**, *277*, 1107–1117.
- (8) (a) Gomez, D.; O'Donohue, M. F.; Wenner, T.; Douarre, C.; Macadré, J.; Koebel, P.; Giraud-Panis, M. J.; Kaplan, H.; Kolkes, A.; Shin-ya, K.; Riou, J. F. The G-quadruplex ligand telomestatin inhibits POT1 binding to telomeric sequences in vitro and induces GFP-POT1 dissociation from telomeres in human cells. *Cancer Res.* **2006**, *66*, 6908–6912. (b) Pagano, B.; Giancola, C. Energetics of quadruplex-drug recognition in anticancer therapy. *Curr. Cancer Drug Targets* **2007**, *7*, 520–540. (c) Salvati, E.; Leonetti, C.; Rizzo, A.; Scarsella, M.; Mottese, M.; Galati, R.; Sperduti, L.; Stevens, M. F.; D'Incalci, M.; Blasco, M.; Chiorino, G.; Bauwens, S.; Horard, B.; Gilson, E.; Stoppacciaro, A.; Zupi, G.; Biroccio, A. Telomere damage induced by the G-quadruplex ligand RHPS4 has an antitumor effect. *J. Clin. Invest.* **2007**, *117*, 3236–3247.
- (9) Burger, A. M.; Dai, F.; Schultes, C. M.; Reszka, A. P.; Moore, M. J.; Double, J. A.; Neidle, S. The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. *Cancer Res.* **2005**, *65*, 1489–1496.
- (10) Phatak, P.; Cookson, J. C.; Dai, F.; Smith, V.; Gartenhaus, R. B.; Stevens, M. F.; Burger, A. M. Telomere uncapping by the G-quadruplex ligand RHPS4 inhibits clonogenic tumour cell growth in vitro and in vivo consistent with a cancer stem cell targeting mechanism. *Br. J. Cancer* **2007**, *96*, 1223–1233.
- (11) Tauchi, T.; Shin-ya, K.; Sashida, G.; Sumi, M.; Okabe, S.; Ohyashiki, J. H.; Ohyashiki, K. Telomerase inhibition with a novel G-quadruplex-interactive agent, telomestatin: in vitro and in vivo studies in acute leukemia. *Oncogene* **2006**, *25*, 5719–5725.
- (12) Cosconati, S.; Marinelli, L.; Trotta, R.; Virno, A.; Mayol, L.; Novellino, E.; Olson, A. J.; Randazzo, A. Tandem application of virtual screening and NMR experiments in the discovery of brand new DNA quadruplex groove binders. *J. Am. Chem. Soc.* **2009**, *131*, 16336–16337.
- (13) (a) Li, Q.; Xiang, J.; Li, X.; Chen, L.; Xu, X.; Tang, Y.; Zhou, Q.; Li, L.; Zhang, H.; Sun, H.; Guan, A.; Yang, Q.; Yang, S.; Xu, G. Stabilizing parallel G-quadruplex DNA by a new class of ligands: two non-planar alkaloids through interaction in lateral grooves. *Biochimie* **2009**, *91*, 811–829. (b) Ranjan, N.; Andreasen, K. F.; Kumar, S.; Hyde-Volpe, D.; Arya, D. P. Aminoglycoside binding to *Oxytricha nova* telomeric DNA. *Biochemistry* **2010**, *49*, 9891–9903. (c) Hamon, F.; Largy, E.; Guédin-Beaupaire, A.; Rouchon-Dagois, M.; Sidibe, A.; Monchaud, D.; Mergny, J. L.; Riou, J. F.; Nguyen, C. H.; Teulade-Fichou, M. P. An acyclic oligoheteroaryle that discriminates strongly between diverse G-quadruplex topologies. *Angew. Chem., Int. Ed.* **2011**, *50*, 8745–8749.
- (14) Trotta, R.; De Tito, S.; Lauri, I.; La Pietra, V.; Marinelli, L.; Cosconati, S.; Martino, L.; Conte, M. R.; Mayol, L.; Novellino, E.; Randazzo, A. A more detailed picture of the interactions between virtual screening-derived hits and the DNA G-quadruplex: NMR, molecular modelling and ITC studies. *Biochimie* **2011**, *93*, 1280–1287.
- (15) (a) Randazzo, A.; Galeone, A.; Mayol, L. ¹H-NMR study of the interaction of distamycin A and netropsin with the parallel stranded tetraplex [d(TGGGGT)]₄. *Chem. Commun.* **2001**, *11*, 1030–1031. (b) Randazzo, A.; Galeone, A.; Esposito, V.; Varra, M.; Mayol, L. Interaction of distamycin A and netropsin with quadruplex and duplex structures: a comparative ¹H-NMR study. *Nucleosides, Nucleotides Nucleic Acids* **2002**, *21*, 535–545. (c) Martino, L.; Virno, A.; Pagano, B.; Virgilio, A.; Di Micco, S.; Galeone, A.; Giancola, C.; Bifulco, G.; Mayol, L.; Randazzo, A. Structural and thermodynamic studies of the interaction of distamycin A with the parallel quadruplex structure [d(TGGGGT)]₄. *J. Am. Chem. Soc.* **2007**, *129*, 16048–16056. (d) Pagano, B.; Virno, A.; Mattia, C. A.; Mayol, L.; Randazzo, A.; Giancola, C. Targeting DNA quadruplexes with distamycin A and its derivatives: an ITC and NMR study. *Biochimie* **2008**, *90*, 1224–12232. (e) Cosconati, S.; Marinelli, L.; Trotta, R.; Virno, A.; De Tito, S.; Romagnoli, R.; Pagano, B.; Limongelli, V.; Giancola, C.; Baraldi, P. G.; Mayol, L.; Novellino, E.; Randazzo, A. Structural and conformational requisites in DNA quadruplex groove binding: another piece to the puzzle. *J. Am. Chem. Soc.* **2010**, *132*, 6425–6433.
- (16) Borges, F.; Roleira, F.; Milhazes, N.; Santana, L.; Uriarte, E. Simple coumarins and analogues in medicinal chemistry: occurrence, synthesis and biological activity. *Curr. Med. Chem.* **2005**, *12*, 887–916.
- (17) Petraccone, L.; Fotticchia, I.; Cummaro, A.; Pagano, B.; Ginnari-Satriani, L.; Haider, S.; Randazzo, A.; Novellino, E.; Neidle, S.; Giancola, C. The triazatruxene derivative azatrux binds to the parallel form of the human telomeric G-quadruplex under molecular crowding conditions: biophysical and molecular modeling studies. *Biochimie* **2011**, *93*, 1318–1327.
- (18) Pagano, B.; Cosconati, S.; Gabelica, V.; Petraccone, L.; De Tito, S.; Marinelli, L.; La Pietra, V.; Di Leva, F. S.; Lauri, I.; Trotta, R.; Novellino, E.; Giancola, C.; Randazzo, A. State-of-the-art methodologies for the discovery and characterization of DNA G-quadruplex binders. *Curr. Pharm. Des.* **2012**, *18*, 1880–1899.
- (19) Horton, D. A.; Bourne, G. T.; Smythe, M. L. The combinatorial synthesis of bicyclic privileged structures or privileged substructures. *Chem. Rev.* **2003**, *103*, 893–930.
- (20) Velazquez Campoy, A.; Freire, E. ITC in the post-genomic era...? Priceless. *Biophys. Chem.* **2005**, *115*, 115–124.
- (21) (a) Pagano, B.; Mattia, C. A.; Virno, A.; Randazzo, A.; Mayol, L.; Giancola, C. Thermodynamic analysis of quadruplex DNA-drug interaction. *Nucleosides, Nucleotides Nucleic Acids* **2007**, *26*, 761–765. (b) Pagano, B.; Mattia, C. A.; Giancola, C. Applications of isothermal titration calorimetry in biophysical studies of G-quadruplexes. *Int. J. Mol. Sci.* **2009**, *10*, 2935–2957.
- (22) Kleit, R. E.; Wemmer, D. E.; Reid, B. R. ¹H NMR studies on the interaction between distamycin A and a symmetrical DNA dodecamer. *Biochemistry* **1986**, *25*, 3296–3303.
- (23) van Steensel, B.; de Lange, T. Control of telomere length by the human telomeric protein TRF1. *Nature* **1997**, *385*, 740–743.
- (24) Takai, H.; Smogorzewska, A.; de Lange, T. DNA damage foci at dysfunctional telomeres. *Curr. Biol.* **2003**, *13*, 1549–1556.
- (25) (a) Yu, H.; Wang, X.; Fu, M.; Ren, J.; Qu, X. Chiral metallo-supramolecular complexes selectively recognize human telomeric G-quadruplex DNA. *Nucleic Acids Res.* **2008**, *36*, 5695–5703. (b) Yu, H.; Zhao, C.; Chen, Y.; Fu, M.; Ren, J.; Qu, X. DNA loop sequence as the determinant for chiral supramolecular compound G-quadruplex selectivity. *J. Med. Chem.* **2010**, *53*, 492–498.
- (26) For example, see the following: (a) D'Onofrio, J.; Petraccone, L.; Martino, L.; Di Fabio, G.; Iadonisi, A.; Balzarini, J.; Giancola, C.; Montesarchio, D. Synthesis, biophysical characterization, and anti-HIV activity of glyco-conjugated G-quadruplex-forming oligonucleotides.

Bioconjugate Chem. **2008**, *19*, 607–616. (b) Di Fabio, G.; D'Onofrio, J.; Chiapparelli, M.; Hoorelbeke, B.; Montesarchio, D.; Balzarini, J.; De Napoli, L. Discovery of novel anti-HIV active G-quadruplex-forming oligonucleotides. *Chem. Commun.* **2011**, *47*, 2363–2365.

(27) Hwang, T. L.; Shaka, A. J. J. Water suppression that works. Excitation sculpting using arbitrary waveforms and pulsed field gradients. *J. Magn. Reson.* **1995**, *A112*, 275–279.

(28) Dalvit, C. Efficient multiple-solvent suppression for the study of the interactions of organic solvents with biomolecules. *J. Biomol. NMR* **1998**, *11*, 437–444.

(29) Cantor, C. R.; Warshaw, M. M.; Shapiro, H. Oligonucleotide interactions. 3. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers* **1970**, *9*, 1059–1077.

(30) Biroccio, A.; Amodei, S.; Benassi, B.; Scarsella, M.; Cianciulli, A.; Mottolese, M.; Del Bufalo, D.; Leonetti, C.; Zupi, G. Reconstitution of hTERT restores tumorigenicity in melanoma-derived c-Myc low-expressing clones. *Oncogene* **2002**, *21*, 3011–3019.

(31) Biroccio, A.; Benassi, B.; Filomeni, G.; Amodei, S.; Marchini, S.; Chiorino, G.; Rotilio, G.; Zupi, G.; Ciriolo, M. R. Glutathione influences c-Myc-induced apoptosis in M14 human melanoma cells. *J. Biol. Chem.* **2002**, *277*, 43763–43770.

(32) Dimri, G. P.; Lee, X.; Basile, G.; Acosta, M.; Scott, G.; Roskelley, C.; Medrano, E. E.; Linskens, M.; Rubelj, L.; Pereira-Smith, O.; Peacocke, M.; Campisi, J. A biomarker that identifies senescent human cells in culture and in aging skin in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9363–9367.